```
        10          20          30          40          50          60          70
CCCTCAGCCC TGCTGCCCAG CACGAGCCTG TGCTCGCCCT GCCCAACGCA GACAGCCAGA CCCAGGGCGG

        80          90         100         110         120         130         140
CCCCTCTGGC GGCTCTGCTC CTCCCGAAGG ATGCTTGGGG AGTGAGGCGA AGCTGGGCGC TCCTCTCCCC

       150         160         170         180         190         200         210
TACAGCAGCC CCCTTCCTCC ATCCCTCTGT TCTCCTGAGC CTTCAGGAGC CTGCACCAGT CCTGCCTGTC

       220         230         240         250         260         270         280
CTTCTACTCA GCTGTTACCC ACTCTGGGAC CAGCAGTCTT TCTGATAACT GGGAGAGGGC AGTAAGGAGG

       290         300         310         320         330         340         350
ACTTCCTGGA GGGGGTGACT GTCCAGAGCC TGGAACTGTG CCCACACCAG AAGCCATCAG CAGCAAGGAC
```

		359			368			377			386			395			404	
ACC	ATG	CGG	CTT	CCG	GGT	GCG	ATG	CCA	GCT	CTG	GCC	CTC	AAA	GGC	GAG	CTG	CTG	
	M	R	L	P	G	A	M	P	A	L	A	L	K	G	E	L	L	17

		413			422			431			440			449			458	
TTG	CTG	TCT—CTC	CTG	TTA	CTT	CTG	GAA	CCA	CAG	ATC	TCT	CAG	GGC	CTG	GTC	GTC		
L	L	S	L	L	L	L	L	E	P	Q	I	S	Q	G	L	V	V	35

		467			476			485			494			503			512	
ACA	CCC	CCG	GGG	CCA	GAG	CTT	GTC	CTC	AAT	GTC	TCC	AGC	ACC	TTC	GTT	CTG	ACC	
T	P	P	G	P	E	L	V	L	N	V	S	S	T	F	V	L	T	53

		521			530			539			548			557			566	
TGC	TCG	GGT	TCA	GCT	CCG	GTG	GTG	TGG	GAA	CGG	ATG	TCC	CAG	GAG	CCC	CCA	CAG	
C	S	G	S	A	P	V	V	W	E	R	M	S	Q	E	P	P	Q	71

		575			584			593			602			611			620	
GAA	ATG	GCC	AAG	GCC	CAG	GAT	GGC	ACC	TTC	TCC	AGC	GTG	CTC	ACA	CTG	ACC	AAC	
E	M	A	K	A	Q	D	G	T	F	S	S	V	L	T	L	T	N	89

		629			638			647			656			665			674	
CTC	ACT	GGG	CTA	GAC	ACG	GGA	GAA	TAC	TTT	TGC	ACC	CAC	AAT	GAC	TCC	CGT	GGA	
L	T	G	L	D	T	G	E	Y	F	C	T	H	N	D	S	R	G	107

		683			692			701			710			719			728	
CTG	GAG	ACC	GAT	GAG	CGG	AAA	CGG	CTC	TAC	ATC	TTT	GTG	CCA	GAT	CCC	ACC	GTG	
L	E	T	D	E	R	K	R	L	Y	I	F	V	P	D	P	T	V	125

		737			746			755			764			773			782	
GGC	TTC	CTC	CCT	AAT	GAT	GCC	GAG	GAA	CTA	TTC	ATC	TTT	CTC	ACG	GAA	ATA	ACT	
G	F	L	P	N	D	A	E	E	L	F	I	F	L	T	E	I	T	143

		791			800			809			818			827			836	
GAG	ATC	ACC	ATT	CCA	TGC	CGA	GTA	ACA	GAC	CCA	CAG	CTG	GTG	GTG	ACA	CTG	CAC	
E	I	T	I	P	C	R	V	T	D	P	Q	L	V	V	T	L	H	161

		845			854			863			872			881			890	
GAG	AAG	AAA	GGG	GAC	GTT	GCA	CTG	CCT	GTC	CCC	TAT	GAT	CAC	CAA	CGT	GGC	TTT	
E	K	K	G	D	V	A	L	P	V	P	Y	D	H	Q	R	G	F	179

		899			908			917			926			935			944	
TCT	GGT	ATC	TTT	GAG	GAC	AGA	AGC	TAC	ATC	TGC	AAA	ACC	ACC	ATT	GGG	GAC	AGG	
S	G	I	F	E	D	R	S	Y	I	C	K	T	T	I	G	D	R	197

```
          953               962              971              980              989              998
GAG GTG GAT TCT GAT GCC TAC TAT GTC TAC AGA CTC CAG GTG TCA TCC ATC AAC
 E   V   D   S   D   A   Y   Y   V   Y   R   L   Q   V   S   S   I   N       215

         1007             1016             1025             1034             1043             1052
GTC TCT GTG AAC GCA GTG CAG ACT GTG GTC CGC CAG GGT GAG AAC ATC ACC CTC
 V   S   V   N   A   V   Q   T   V   V   R   Q   G   E   N   I   T   L       233

         1061             1070             1079             1088             1097             1106
ATG TGC ATT GTG ATC GGG AAT GAG GTG GTC AAC TTC GAG TGG ACA TAC CCC CGC
 M   C   I   V   I   G   N   E   V   V   N   F   E   W   T   Y   P   R       251

         1115             1124             1133             1142             1151             1160
AAA GAA AGT GGG CGG CTG GTG GAG CCG GTG ACT GAC TTC CTC TTG GAT ATG CCT
 K   E   S   G   R   L   V   E   P   V   T   D   F   L   L   D   M   P       269

         1169             1178             1187             1196             1205             1214
TAC CAC ATC CGC TCC ATC CTG CAC ATC CCC AGT GCC GAG TTA GAA GAC TCG GGG
 Y   H   I   R   S   I   L   H   I   P   S   A   E   L   E   D   S   G       287

         1223             1232             1241             1250             1259             1268
ACC TAC ACC TGC AAT GTG ACG GAG AGT GTG AAT GAC CAT CAG GAT GAA AAG GCC
 T   Y   T   C   N   V   T   E   S   V   N   D   H   Q   D   E   K   A       305

         1277             1286             1295             1304             1313             1322
ATC AAC ATC ACC GTG GTT GAG AGC GGC TAC GTG CGG CTC CTG GGA GAG GTG GGC
 I   N   I   T   V   V   E   S   G   Y   V   R   L   L   G   E   V   G       323

         1331             1340             1349             1358             1367             1376
ACA CTA CAA TTT GCT GAG CTG CAT CGG AGC CGG ACA CTG CAG GTA GTG TTC GAG
 T   L   Q   F   A   E   L   H   R   S   R   T   L   Q   V   V   F   E       341

         1385             1394             1403             1412             1421             1430
GCC TAC CCA CCG CCC ACT GTC CTG TGG TTC AAA GAC AAC CGC ACC CTG GGC GAC
 A   Y   P   P   P   T   V   L   W   F   K   D   N   R   T   L   G   D       359

         1439             1448             1457             1466             1475             1484
TCC AGC GCT GGC GAA ATC GCC CTG TCC ACG CGC AAC GTG TCG GAG ACC CGG TAT
 S   S   A   G   E   I   A   L   S   T   R   N   V   S   E   T   R   Y       377

         1493             1502             1511             1520             1529             1538
GTG TCA GAG CTG ACA CTG GTT CGC GTG AAG GTG GCA GAG GCT GGC CAC TAC ACC
 V   S   E   L   T   L   V   R   V   K   V   A   E   A   G   H   Y   T       391

         1547             1556             1565             1574             1583             1592
ATG CGG GCC TTC CAT GAG GAT GCT GAG GTC CAG CTC TCC TTC CAG CTA CAG ATC
 M   R   A   F   H   E   D   A   E   V   Q   L   S   F   Q   L   Q   I       413

         1601             1610             1619             1628             1637             1646
AAT GTC CCT GTC CGA GTG CTG GAG CTA AGT GAG AGC CAC CCT GAC AGT GGG GAA
 N   V   P   V   R   V   L   E   L   S   E   S   H   P   D   S   G   E       431

         1655             1664             1673             1682             1691             1700
CAG ACA GTC CGC TGT CGT GGC CGG GGC ATG CCC CAG CCG AAC ATC ATC TGG TCT
 Q   T   V   R   C   R   G   R   G   M   P   Q   P   N   I   I   W   S       449

         1709             1718             1727             1736             1745             1754
GCC TGC AGA GAC CTC AAA AGG TGT CCA CGT GAG CTG CCG CCC ACG CTG CTG GGG
 A   C   R   D   L   K   R   C   P   R   E   L   P   P   T   L   L   G       467
```

```
        1763        1772           1781          1790          1799          1808
AAC AGT TCC GAA GAG GAG AGC CAG CTG GAG ACT AAC GTG ACG TAC TGG GAG GAG
 N   S   S   E   E   E   S   Q   L   E   T   N   V   T   Y   W   E   E      485

        1817        1826          1835           1844          1853          1862
GAG CAG GAG TTT GAG GTG GTG AGC ACA CTG CGT CTG CAG CAC GTG GAT CGG CCA
 E   Q   E   F   E   V   V   S   T   L   R   L   Q   H   V   D   R   P      503

        1871        1880          1889          1898          1907          1916
CTG TCG GTG CGC TGC ACG CTG CGC AAC GCT GTG GGC CAG GAC ACG CAG GAG GTC
 L   S   V   R   C   T   L   R   N   A   V   G   Q   D   T   Q   E   V      521

        1925        1934          1943          1952          1961          1970
ATC GTG GTG CCA CAC TCC TTG CCC TTT AAG GTG GTG GTG ATC TCA GCC ATC CTG
 I   V   V   P   H   S   L   P   F   K   V   V   V   I   S   A   I   L      539

        1979        1988          1997          2006          2015          2024
GCC CTG GTG GTG CTC ACC ATC ATC TCC CTT ATC ATC CTC ATC ATG CTT TGG CAG
 A   L   V   V   L   T   I   I   S   L   I   I   L   I   M   L   W   Q      557

        2033        2042          2051          2060          2069          2078
AAG AAG CCA CGT TAC GAG ATC CGA TGG AAG GTG ATT GAG TCT GTG AGC TCT GAC
 K   K   P   R   Y   E   I   R   W   K   V   I   E   S   V   S   S   D      575

        2087        2096          2105          2114          2123          2132
GGC CAT GAG TAC ATC TAC GTG GAC CCC ATG CAG CTG CCC TAT GAC TCC ACG TGG
 G   H   E   Y   I   Y   V   D   P   M   Q   L   P   Y   D   S   T   W      593

        2141        2150          2159          2168          2177          2186
GAG CTG CCG CGG GAC CAG CTT GTG CTG GGA CGC ACC CTC GGC TCT GGG GCC TTT
 E   L   P   R   D   Q   L   V   L   G   R   T   L   G   S   G   A   F      611

        2195        2204          2213          2222          2231          2240
GGG CAG GTG GTG GAG GCC ACG GCT CAT GGC CTG AGC CAT TCT CAG GCC ACG ATG
 G   Q   V   V   E   A   T   A   H   G   L   S   H   S   Q   A   T   M      629

        2249        2258          2267          2276          2285          2294
AAA GTG GCC GTC AAG ATG CTT AAA TCC ACA GCC CGC AGC AGT GAG AAG CAA GCC
 K   V   A   V   K   M   L   K   S   T   A   R   S   S   E   K   Q   A      647

        2303        2312          2321          2330          2339          2348
CTT ATG TCG GAG CTG AAG ATC ATG AGT CAC CTT GGG CCC CAC CTG AAC GTG GTC
 L   M   S   E   L   K   I   M   S   H   L   G   P   H   L   N   V   V      665

        2357        2366          2375          2384          2393          2402
AAC CTG TTG GGG GCC TGC ACC AAA GGA GGA CCC ATC TAT ATC ATC ACT GAG TAC
 I   L   L   G   A   C   T   K   G   G   P   I   Y   I   I   T   E   Y      683

        2411        2420          2429          2438          2447          2456
CGC CGC TAC GGA GAC CTG GTG GAC TAC CTG CAC CGC AAC AAA CAC ACC TTC CTG
 R   R   Y   G   D   L   V   D   Y   L   H   R   N   K   H   T   F   L      701

        2465        2474          2483          2492          2501          2510
CAG CAC CAC TCC GAC AAG CGC CGC CCG CCC AGC GCG GAG CTC TAC AGC AAT GCT
 H   H   S   D   K   R   R   P   P   S   A   E   L   Y   S   N   A        719

        2519        2528          2537          2546          2555          2564
TTG CCC GTT GGG CTC CCC CTG CCC AGC CAT GTG TCC TTG ACC GGG GAG AGC GAC
 P   V   G   L   P   L   P   S   H   V   S   L   T   G   E   S   D        737
```

```
     2573           2582              2591              2600              2609              2618
GGT GGC TAC ATG GAC ATG AGC AAG GAC GAG TCG GTG GAC TAT GTG CCC ATG CTG
 G   G   Y   M   D   M   S   K   D   E   S   V   D   Y   V   P   M   L      755

     2627           2636              2645              2654              2663              2672
GAC ATG AAA GGA GAC GTC AAA TAT GCA GAC ATC GAG TCC TCC AAC TAC ATG GCC
 D   M   K   G   D   V   K   Y   A   D   I   E   S   S   N   Y   M   A      773

     2681           2690              2699              2708              2717              2726
CCT TAC GAT AAC TAC GTT CCC TCT GCC CCT GAG AGG ACC TGC CGA GCA ACT TTG
 P   Y   D   N   Y   V   P   S   A   P   E   R   T   C   R   A   T   L      791

     2735           2744              2753              2762              2771              2780
ATC AAC GAG TCT CCA GTG CTA AGC TAC ATG GAC CTC GTG GGC TTC AGC TAC CAG
 I   N   E   S   P   V   L   S   Y   M   D   L   V   G   F   S   Y   Q      809

     2789           2798              2807              2816              2825              2834
GTG GCC AAT GGC ATG GAG TTT CTG GCC TCC AAG AAC TGC GTC CAC AGA GAC CTG
 V   A   N   G   M   E   F   L   A   S   K   N   C   V   H   R   D   L      827

     2843           2852              2861              2870              2879              2888
GCG GCT AGG AAC GTG CTC ATC TGT GAA GGC AAG CTG GTC AAG ATC TGT GAC TTT
 A   A   R   N   V   L   I   C   E   G   K   L   V   K   I   C   D   F      845

     2897           2906              2915              2924              2933              2942
GGC CTG GCT CGA GAC ATC ATG CGG GAC TCG AAT TAC ATC TCC AAA GGC AGC ACC
 G   L   A   R   D   I   M   R   D   S   N   Y   I   S   K   G   S   T      863

     2951           2960              2969              2978              2987              2996
TTT TTG CCT TTA AAG TGG ATG GCT CCG GAG AGC ATC TTC AAC AGC CTC TAC ACC
 F   L   P   L   K   W   M   A   P   E   S   I   F   N   S   L   Y   T      881

     3005           3014              3023              3032              3041              3050
ACC CTG AGC GAC GTG TGG TCC TTC GGG ATC CTG CTC TGG GAG ATC TTC ACC TTG
 T   L   S   D   V   W   S   F   G   I   L   L   W   E   I   F   T   L      899

     3059           3068              3077              3086              3095              3104
GGT GGC ACC CCT TAC CCA GAG CTG CCC ATG AAC GAG CAG TTC TAC AAT GCC ATC
 G   G   T   P   Y   P   E   L   P   M   N   E   Q   F   Y   N   A   I      917

     3113           3122              3131              3140              3149              3158
AAA CGG GGT TAC CGC ATG GCC CAG CCT GCC CAT GCC TCC GAC GAG ATC TAT GAG
 K   R   G   Y   R   M   A   Q   P   A   H   A   S   D   E   I   Y   E      935

     3167           3176              3185              3194              3203              3212
ATC ATG CAG AAG TGC TGG GAA GAG AAG TTT GAG ATT CGG CCC CCC TTC TCC CAG
 I   M   Q   K   C   W   E   E   K   F   E   I   R   P   P   F   S   Q      953

     3221           3230              3239              3248              3257              3266
CTG GTG CTG CTT CTC GAG AGA CTG TTG GGC GAA GGT TAC AAA AAG AAG TAC CAG
 V   L   L   L   E   R   L   L   G   E   G   Y   K   K   K   Y   Q          971

     3275           3284              3293              3302              3311              3320
CAG GTG GAT GAG GAG TTT CTG AGG AGT GAC CAC CCA GCC ATC CTT CGG TCC CAG
 V   D   E   E   F   L   R   S   D   H   P   A   I   L   R   S   Q          989

     3329           3338              3347              3356              3365              3374
CCC CGC TTG CCT GGG TTC CAT GGC CTC CGA TCT CCC CTG GAC ACC AGC TCC GTC
 R   L   P   G   F   H   G   L   R   S   P   L   D   T   S   S   V        1007
```

Figure 1 continued

```
      3383          3392          3401          3410          3419          3428
CTC TAT ACT GCC GTG CAG CCC AAT GAG GGT GAC AAC GAC TAT ATC ATC CCC CTG
 L   Y   T   A   V   Q   P   N   E   G   D   N   D   Y   I   I   P   L    1025

      3437          3446          3455          3464          3473          3482
CCT GAC CCC AAA CCC GAG GTT GCT GAC GAG GGC CCA CTG GAG GGT TCC CCC AGC
 P   D   P   K   P   E   V   A   D   E   G   P   L   E   G   S   P   S    1043

      3491          3500          3509          3518          3527          3536
CTA GCC AGC TCC ACC CTG AAT GAA GTC AAC ACC TCC TCA ACC ATC TCC TGT GAC
 L   A   S   S   T   L   N   E   V   N   T   S   S   T   I   S   C   D    1061

      3545          3554          3563          3572          3581          3590
AGC CCC CTG GAG CCC CAG GAC GAA CCA GAG CCA GAG CCC CAG CTT GAG CTC CAG
 S   P   L   E   P   Q   D   E   P   E   P   E   P   Q   L   E   L   Q    1079

      3599          3608          3617          3626          3635          3644
GTG GAG CCG GAG CCA GAG CTG GAA CAG TTG CCG GAT TCG GGG TGC CCT GCG CCT
 V   E   P   E   P   E   L   E   Q   L   P   D   S   G   C   P   A   P    1097

      3653          3662          3671          3684          3694          3704
CGG GCG GAA GCA GAG GAT AGC TTC CTG TAG GGGGCTGGCC CCTACCCTGC CCTGCCTGAA
 R   A   E   A   E   D   S   F   L   .                                    1106

      3714          3724          3734          3744          3754          3764          3774
GCTCCCCCCC TGCCAGCACC CAGCATCTCC TGGCCTGGCC TGACCGGGCT TCCTGTCAGC CAGGCTGCCC

      3784          3794          3804          3814          3824          3834          3844
TTATCAGCTG TCCCCTTCTG GAAGCTTTCT GCTCCTGACG TGTTGTGCCC CAAACCCTGG GGCTGGCTTA

      3854          3864          3874          3884          3894          3904          3914
GGAGGCAAGA AAACTGCAGG GGCCGTGACC AGCCCTCTGC CTCCAGGGAG GCCAACTGAC TCTGAGCCAG

      3924          3934          3944          3954          3964          3974          3984
GGTTCCCCCA GGGAACTCAG TTTTCCCATA TGTAAGATGG GAAAGTTAGG CTTGATGACC CAGAATCTAG

      3994          4004          4014          4024          4034          4044          4054
GATTCTCTCC CTGGCTGACA GGTGGGGAGA CCGAATCCCT CCCTGGGAAG ATTCTTGGAG TTACTGAGGT

      4064          4074          4084          4094          4104          4114          4124
GGTAAATTAA CTTTTTTCTG TTCAGCCAGC TACCCCTCAA GGAATCATAG CTCTCTCCTC GCACTTTTTA

      4134          4144          4154          4164          4174          4184          4194
TCCACCCAGG AGCTAGGGAA GAGACCCTAG CCTCCCTGGC TGCTGGCTGA GCTAGGGCCT AGCCTTGAGC

      4204          4214          4224          4234          4244          4254          4264
AGTGTTGCCT CATCCAGAAG AAAGCCAGTC TCCTCCCTAT GATGCCAGTC CCTGCGTTCC CTGGCCCGAG

      4274          4284          4294          4304          4314          4324          4334
CTGGTCTGGG GCCATTAGGC AGCCTAATTA ATGCTGGAGG CTGAGCCAAG TACAGGACAC CCCCAGCCTG

      4344          4354          4364          4374          4384          4394          4404
CAGCCCTTGC CCAGGGCACT TGGAGCACAC GCAGCCATAG CAAGTGCCTG TGTCCCTGTC CTTCAGGCCC

      4414          4424          4434          4444          4454          4464          4474
ATCAGTCCTG GGGCTTTTTC TTTATCACCC TCAGTCTTAA TCCATCCACC AGAGTCTAGA AGGCCAGACG

      4484          4494          4504          4514          4524          4534          4544
GGCCCCGCAT CTGTGATGAG AATGTAAATG TGCCAGTGTG GAGTGGCCAC GTGTGTGTGC CAGTATATGG
```
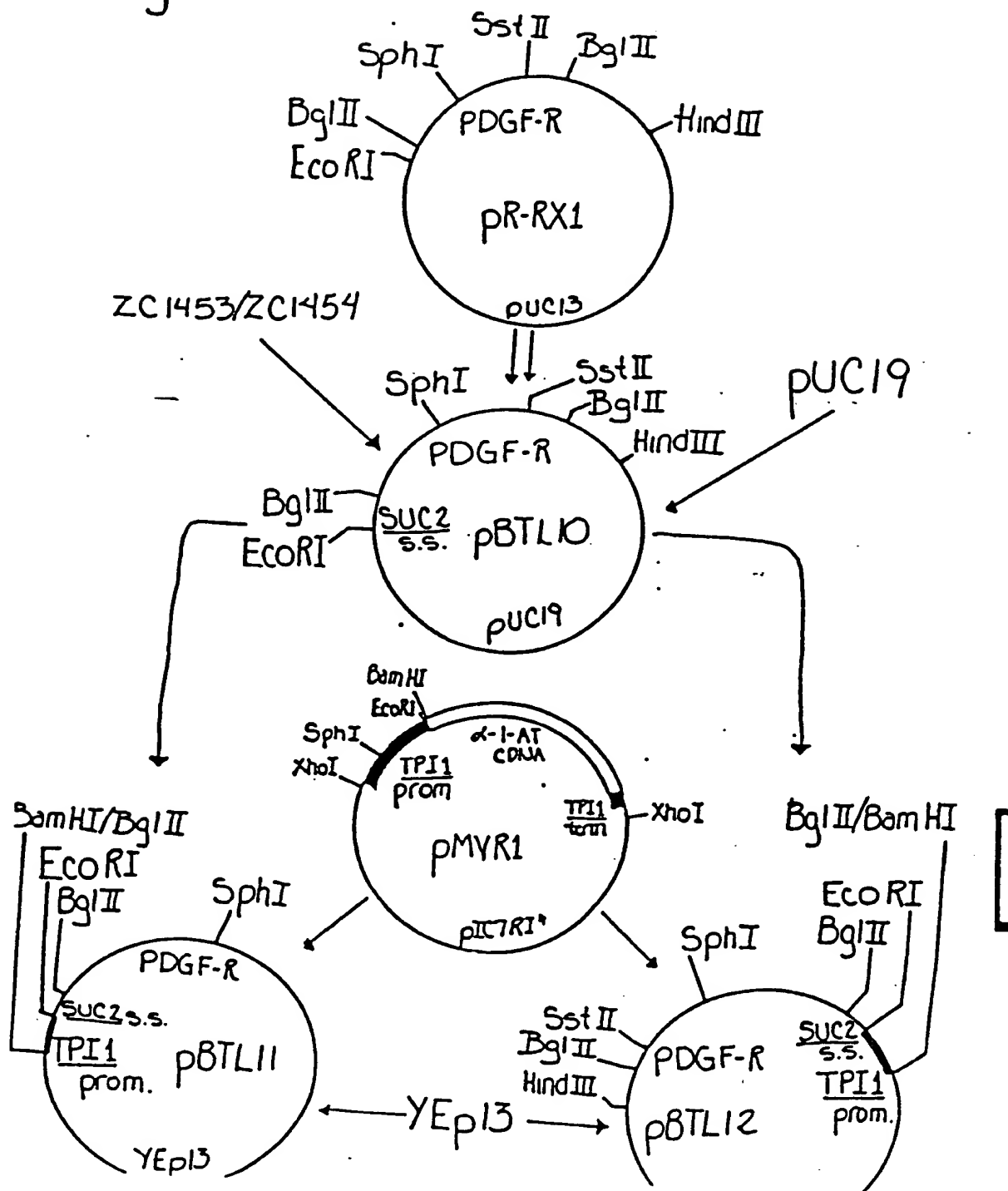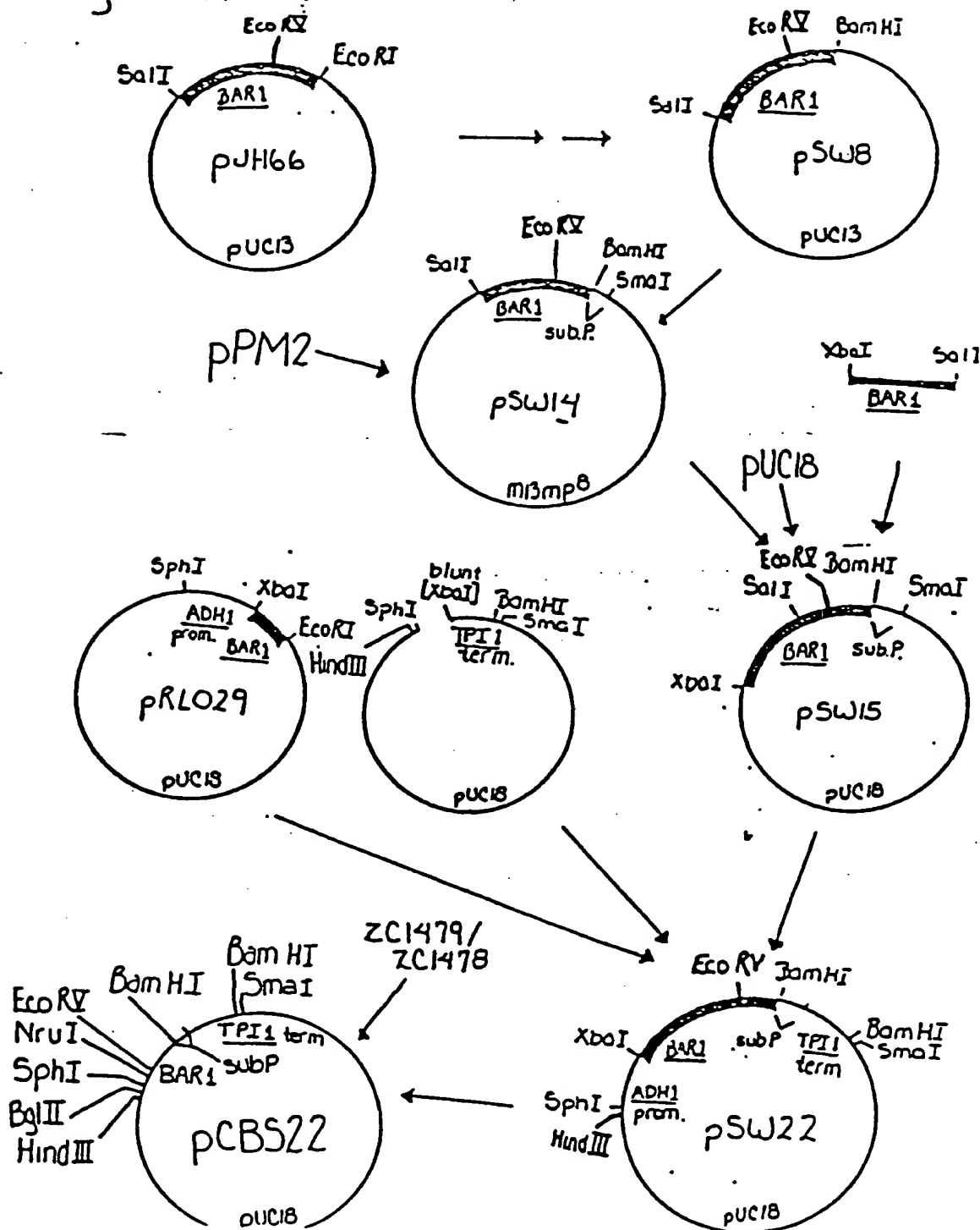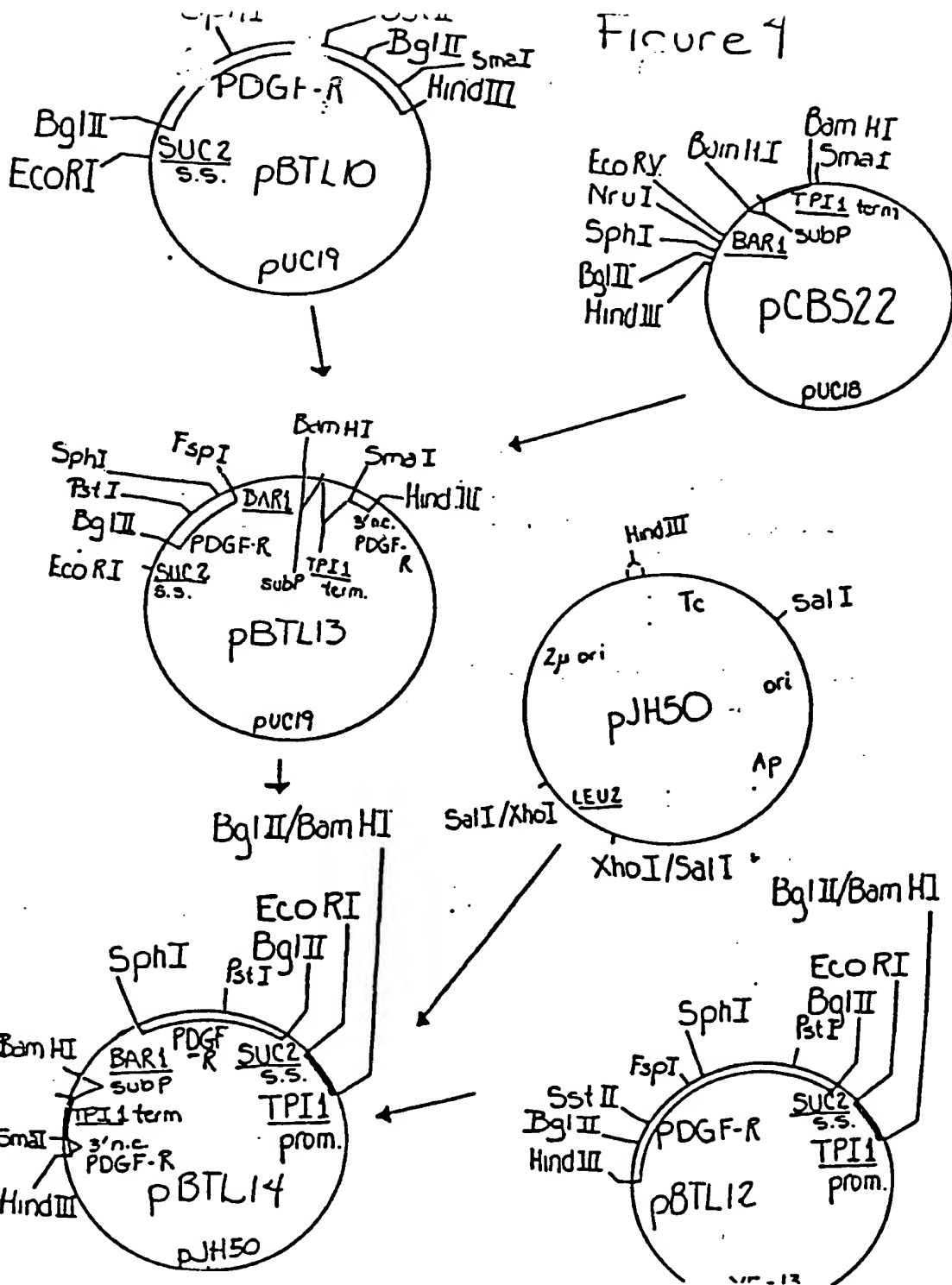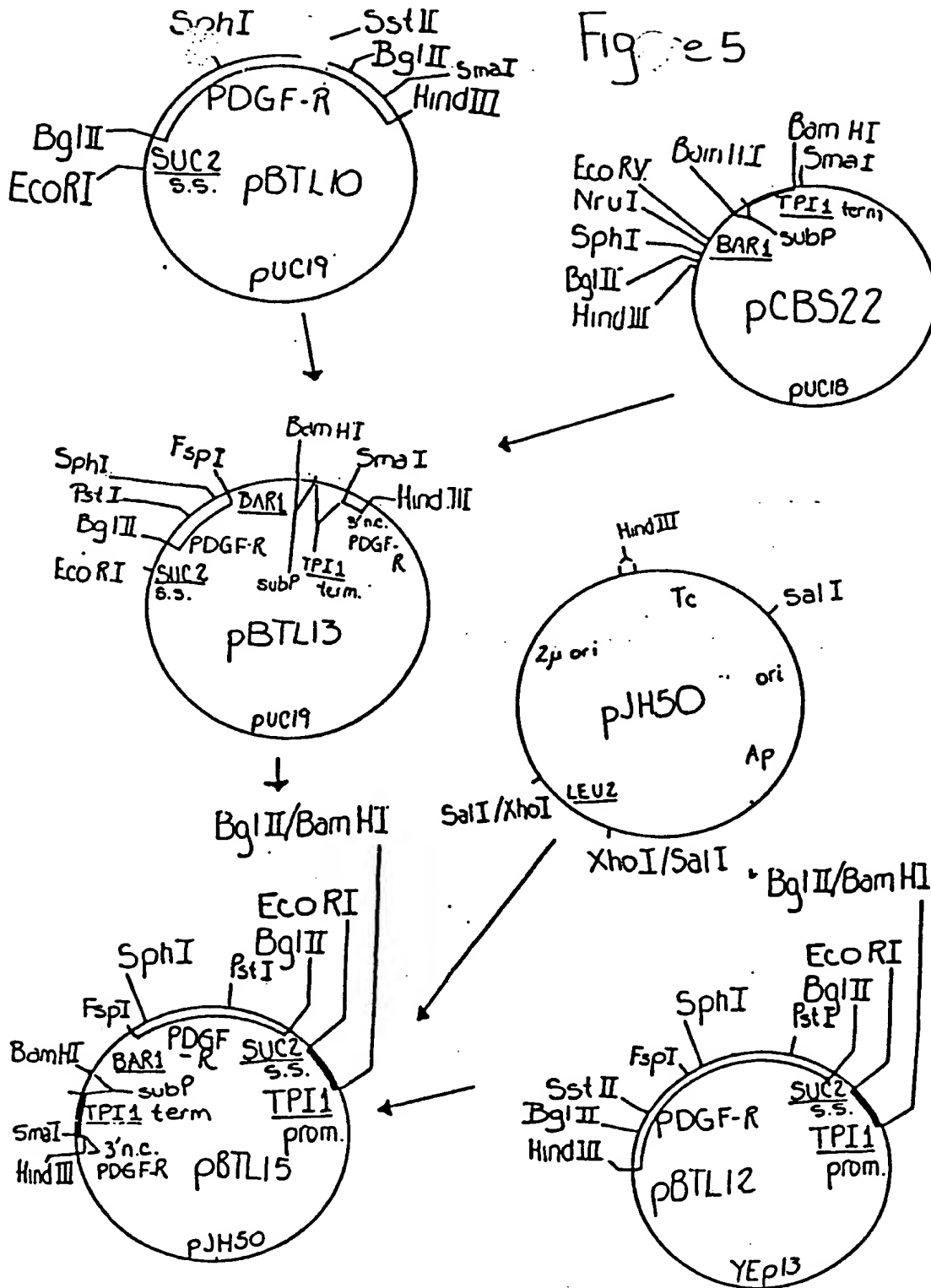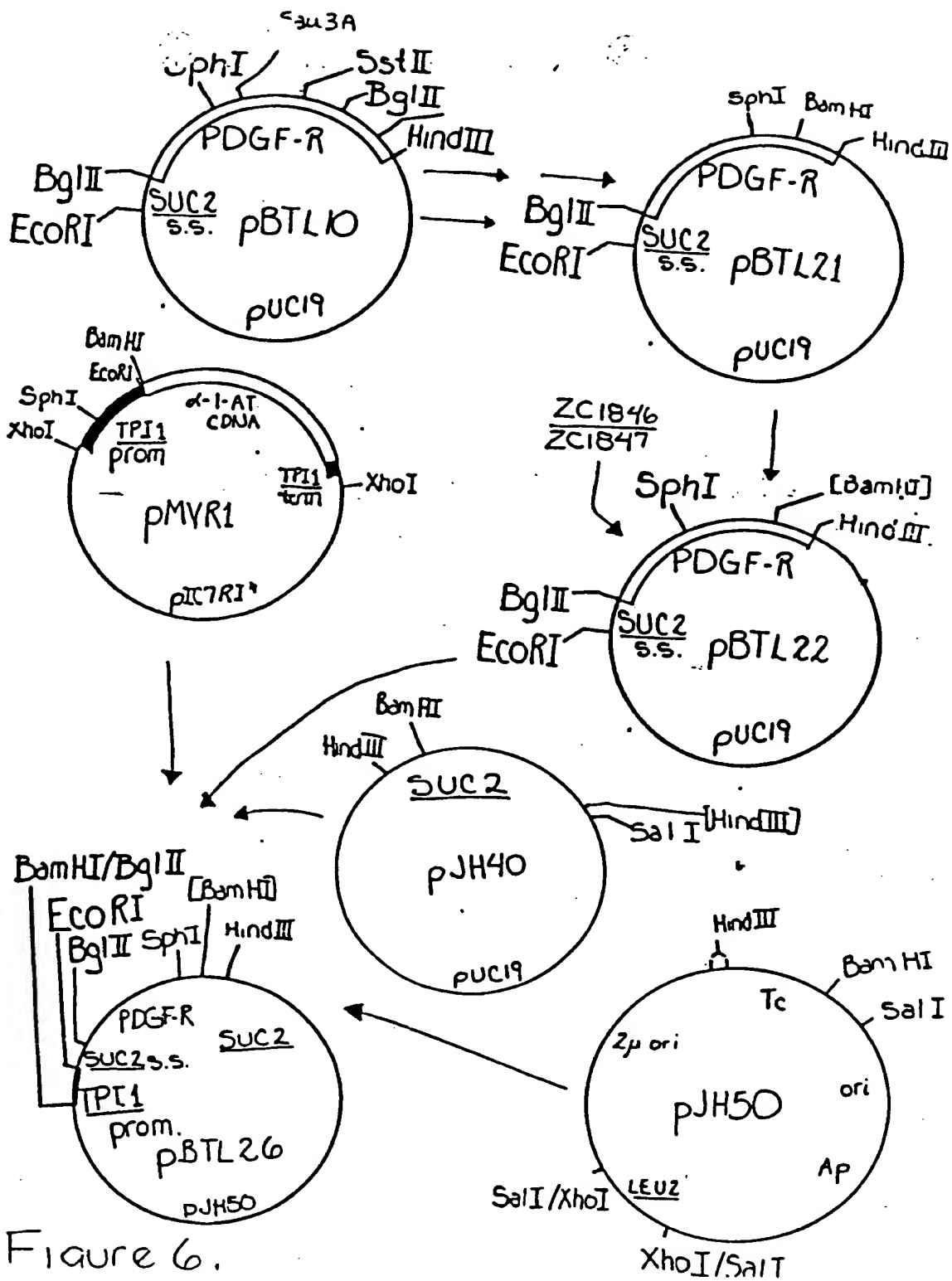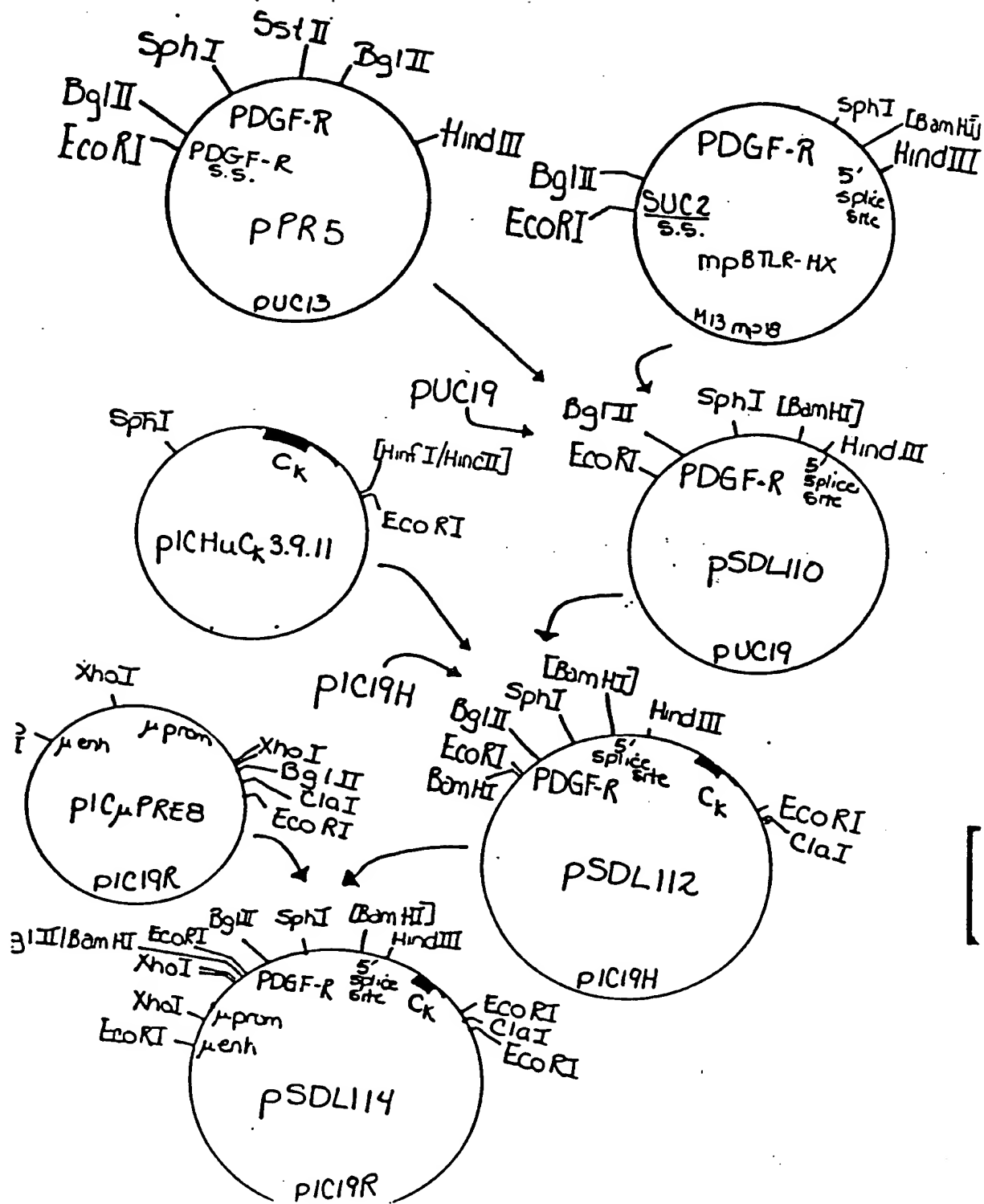
Figure 2
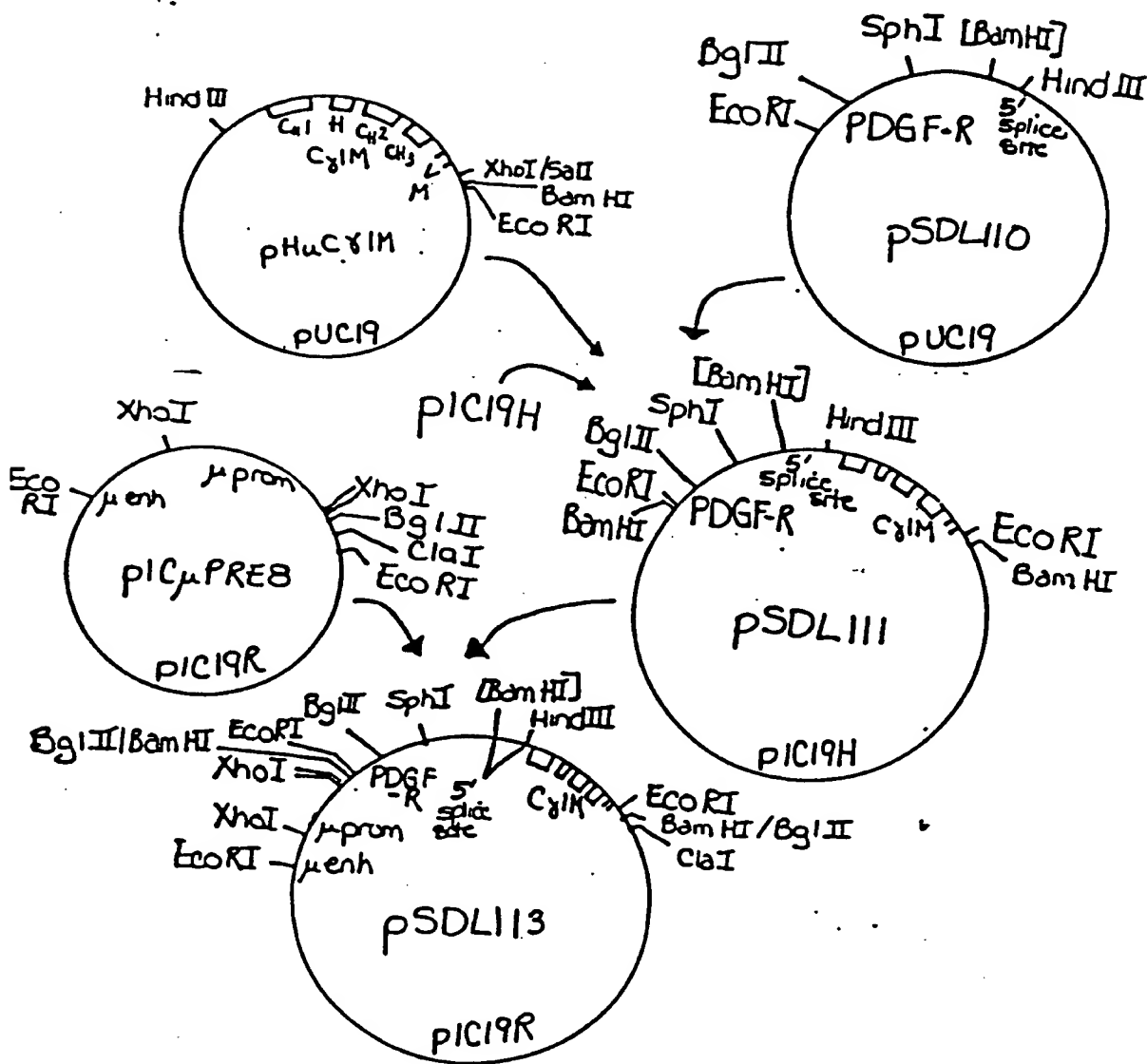
Figure 3.

Figure 4
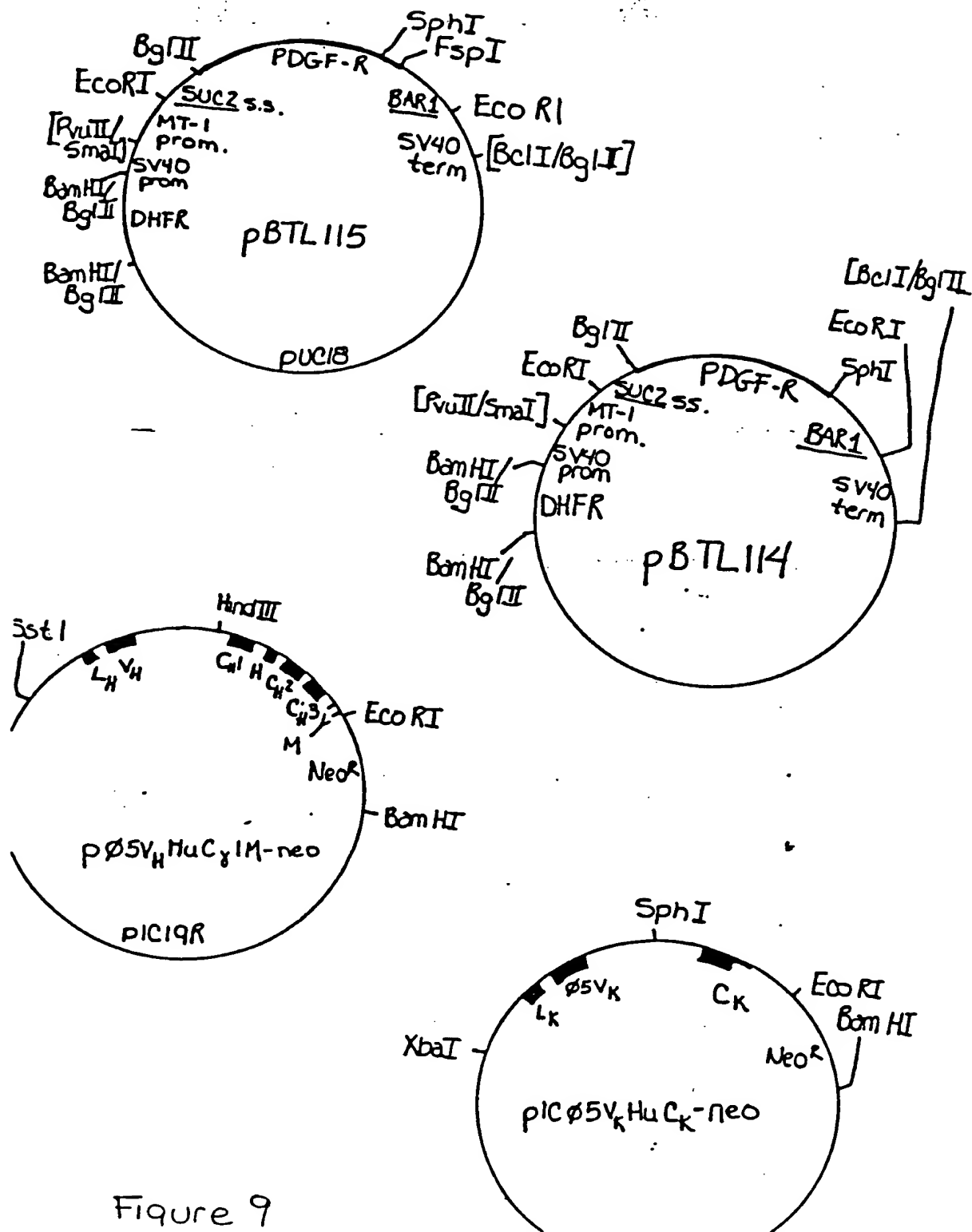
Figure 5

Figure 6.
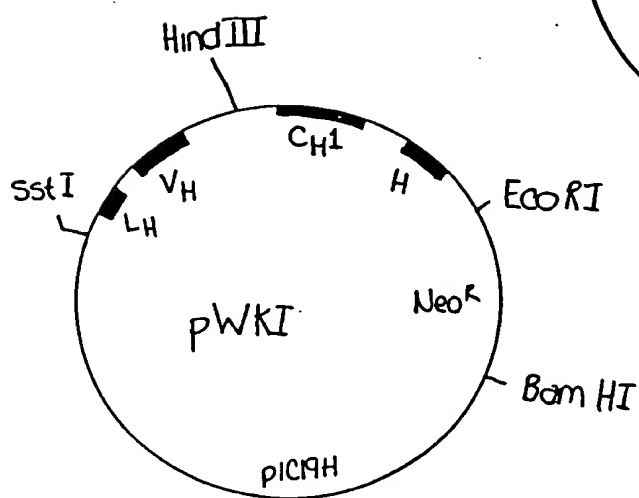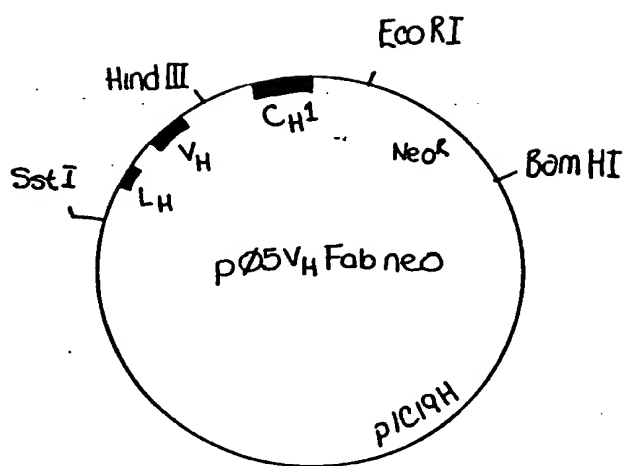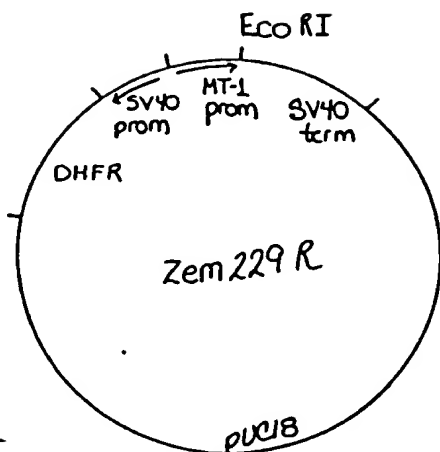
Figure 7

Figure 8

Figure 9

Figure 10

FIGURE 11

```
  1 GCCCTGGGGACGGACCGTGGGCGGCGCGCAGCGGCGGGACGCCGTTTTGGGGACGTGGTGGCCAGCGCCT
 70 TCCTGCAGACCCACAGGGAAGTACTCCCTTTGACCTCCGGGGAGCTGCGACCAGGTTATACGTTGCTGG

139 TGGAAAAGTGACAATTCTAGGAAAAGAGCTAAAAGCCGGATCGGTGACCGAAAGTTTCCCAGAGCTATG
                                                                        M
                                                                        1

208 GGGACTTCCCATCCGGCGTTCCTGGTCTTAGGCTGTCTTCTCACAGGGCTGAGCCTAATCCTCTGCCAG
     G  T  S  H  P  A  F  L  V  L  G  C  L  L  T  G  L  S  L  I  L  C  Q

277 CTTTCATTACCCTCTATCCTTCCAAATGAAAATGAAAAGGTTGTGCAGCTGAATTCATCCTTTTCTCTG
     L  S  L  P  S  I  L  P  N  E  N  E  K  V  V  Q  L  N  S  S  F  S  L

346 AGATGCTTTGGGGAGAGTGAAGTGAGCTGGCAGTACCCCATGTCTGAAGAAGAGAGCTCCGATGTGGAA
     R  C  F  G  E  S  E  V  S  W  Q  Y  P  M  S  E  E  E  S  S  D  V  E

415 ATCAGAAATGAAGAAAACAACAGCGGCCTTTTTGTGACGGTCTTGGAAGTGAGCAGTGCCTCGGCGGCC
     I  R  N  E  E  N  N  S  G  L  F  V  T  V  L  E  V  S  S  A  S  A  A

484 CACACAGGGTTGTACACTTGCTATTACAACCACACTCAGACAGAAGAGAATGAGCTTGAAGGCAGGCAC
     H  T  G  L  Y  T  C  Y  Y  N  H  T  Q  T  E  E  N  E  L  E  G  R  H

553 ATTTACATCTATGTGCCAGACCCAGATGTAGCCTTTGTACCTCTAGGAATGACGGATTATTTAGTCATC
     I  Y  I  Y  V  P  D  P  D  V  A  F  V  P  L  G  M  T  D  Y  L  V  I

622 GTGGAGGATGATGATTCTGCCATTATACCTTGTCGCACAACTGATCCCGAGACTCCTGTAACCTTACAC
     V  E  D  D  D  S  A  I  I  P  C  R  T  T  D  P  E  T  P  V  T  L  H

691 AACAGTGAGGGGGTGGTACCTGCCTCCTACGACAGCAGACAGGGCTTTAATGGGACCTTCACTGTAGGG
     N  S  E  G  V  V  P  A  S  Y  D  S  R  Q  G  F  N  G  T  F  T  V  G

760 CCCTATATCTGTGAGGCCACCGTCAAAGGAAAGAAGTTCCAGACCATCCCATTTAATGTTTATGCTTTA
     P  Y  I  C  E  A  T  V  K  G  K  K  F  Q  T  I  P  F  N  V  Y  A  L

829 AAAGCAACATCAGAGCTGGATCTAGAAATGGAAGCTCTTAAAACCGTGTATAAGTCAGGGGAAACGATT
     K  A  T  S  E  L  D  L  E  M  E  A  L  K  T  V  Y  K  S  G  E  T  I

898 GTGGTCACCTGTGCTGTTTTTAACAATGAGGTGGTTGACCTTCAATGGACTTACCCTGGAGAAGTGAAA
     V  V  T  C  A  V  F  N  N  E  V  V  D  L  Q  W  T  Y  P  G  E  V  K
```

Figure 11 continued

967 GGCAAAGGCATCACAATACTGGAAGAAATCAAAGTCCCATCCATCAAATTGGTGTACACTTTGACGGTC
    G K G I T I L E E I K V P S I K L V Y T L T V

1036 CCCGAGGCCACGGTGAAAGACAGTGGAGATTACGAATGTGCTGCCCGCCAGGCTACCAGGGAGGTCAAA
    P E A T V K D S G D Y E C A A R Q A T R E V K

1105 GAAATGAAGAAAGTCACTATTTCTGTCCATGAGAAAGGTTTCATTGAAATCAAACCCACCTTCAGCCAG
    E M K K V T I S V H E K G F I E I K P T F S Q

1174 TTGGAAGCTGTCAACCTGCATGAAGTCAAACATTTTGTTGTAGAGGTGCGGGCCTACCCACCTCCCAGG
    L E A V N L H E V K H F V V E V R A Y P P P R

1243 ATATCCTGGCTGAAAAACAATCTGACTCTGATTGAAAATCTCACTGAGATCACCACTGATGTGGAAAAG
    I S W L K N N L T L I E N L T E I T T D V E K

1312 ATTCAGGAAATAAGGTATCGAAGCAAATTAAAGCTGATCCGTGCTAAGGAAGAAGACAGTGGCCATTAT
    I Q E I R Y R S K L K L I R A K E E D S G H Y

1381 ACTATTGTAGCTCAAAATGAAGATGCTGTGAAGAGCTATACTTTTGAACTGTTAACTCAAGTTCCTTCA
    T I V A Q N E D A V K S Y T F E L L T Q V P S

1450 TCCATTCTGGACTTGGTCGATGATCACCATGGCTCAACTGGGGGACAGACGGTGAGGTGCACAGCTGAA
    S I L D L V D D H H G S T G G Q T V R C T A E

1519 GGCACGCCGCTTCCTGATATTGAGTGGATGATATGCAAAGATATTAAGAAATGTAATAATGAAACTTCC
    G T P L P D I E W M I C K D I K K C N N E T S

1588 TGGACTATTTTGGCCAACAATGTCTCAAACATCATCACGGAGATCCACTCCCGAGACAGGAGTACCGTG
    W T I L A N N V S N I I T E I H S R D R S T V

1657 GAGGGCCGTGTGACTTTCGCCAAAGTGGAGGAGACCATCGCCGTGCGATGCCTGGCTAAGAATCTCCTT
    E G R V T F A K V E E T I A V R C L A K N L L

1726 GGAGCTGAGAACCGAGAGCTGAAGCTGGTGGCTCCCACCCTGCGTTCTGAACTCACGGTGGCTGCTGCA
    G A E N R E L K L V A P T L R S E L T V A A A

1795 GTCCTGGTGCTGTTGGTGATTGTGATCATCTCACTTATTGTCCTGGTTGTCATTTGGAAACAGAAACCG
    V L V L L V I V I I S L I V L V V I W K Q K P

Figure 11 continued

1864 AGGTATGAAATTCGCTGGAGGGTCATTGAATCAATCAGCCCGGATGGACATGAATATATTTATGTGGAC
     R Y E I R W R V I E S I S P D G H E Y I Y V D

1933 CCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCCAAGAGATGGACTAGTGCTTGGTCGGGTCTTG
     P M Q L P Y D S R W E F P R D G L V L G R V L

2002 GGGTCTGGAGCCGTTTGGGAAGGTGGTTGAAGGAACAGCCTATGGATTAAGCCGGTCCCAACCTGTCATG
     G S G A F G K V V E G T A Y G L S R S Q P V M

2071 AAAGTTGCAGTGAAGATGCTAAAACCCACGGCCAGATCCAGTGAAAAACAAGCTCTCATGTCTGAACTG
     K V A V K M L K P T A R S S E K Q A L M S E L

2140 AAGATAATGACTCACCTGGGGCCACATTTGAACATTGTAAACTTGCTGGGAGCCTGCACCAAGTCAGGC
     K I M T H L G P H L N I V N L L G A C T K S G

2209 CCCATTTACATCATCACAGAGTATTGCTTCTATGGAGATTTGGTCAACTATTTGCATAAGAATAGGGAT
     P I Y I I T E Y C F Y G D L V N Y L H K N R D

2278 AGCTTCCTGAGCCACCACCCAGAGAAGCCAAAGAAAGAGCTGGATATCTTTGGATTGAACCCTGCTGAT
     S F L S H H P E K P K K E L D I F G L N P A D

2347 GAAAGCACACGGAGCTATGTTATTTTATCTTTTGAAAACAATGGTGACTACATGGACATGAAGCAGGCT
     E S T R S Y V I L S F E N N G D Y M D M K Q A

2416 GATACTACACAGTATGTCCCCATGCTAGAAAGGAAAGAGGTTTCTAAATATTCCGACATCCAGAGATCA
     D T T Q Y V P M L E R K E V S K Y S D I Q R S

2485 CTCTATGATCGTCCAGCCTCATATAAGAAGAAATCTATGTTAGACTCAGAAGTCAAAAACCTCCTTTCA
     L Y D R P A S Y K K K S M L D S E V K N L L S

2554 GATGATAACTCAGAAGGCCTTACTTTATTGGATTTGTTGAGCTTCACCTATCAAGTTGCCCGAGGAATG
     D D N S E G L T L L D L L S F T Y Q V A R G M

2623 GAGTTTTTGGCTTCAAAAAAATTGTGTCCACCGTGATCTGGCTGCTCGCAACGTCCTCCTGGCACAAGGA
     E F L A S K N C V H R D L A A R N V L L A Q G

2692 AAAATTGTGAAGATCTGTGACTTTGGCCTGGCCAGAGACATCATGCATGATTQGAACTATGTGTCGAAA
     K I V K I C D F G L A R D I M H D S N Y V S K

2761 GGCAGTACCTTTCTGCCCGTGAAGTGGATGGCTCCTGAGAGCATCTTTGACAACCTCTACACCACACTG
     G S T F L P V K W M A P E S I F D N L Y T T L

2830 AGTGATGTCTGGTCTTATGGCATTCTGCTCTGGGAGATCTTTTCCCTTGGTGGCACCCCTTACCCCGGC
     S D V W S Y G I L L W E I F S L G G T P Y P G

2899 ATGATGGTGGATTCTACTTTCTACAATAAGATCAAGAGTGGGTACCGGATGGCCAAGCCTGACCACGCT
     M M V D S T F Y N K I K S G Y R M A K P D H A

2968 ACCAGTGAAGTCTACGAGATCATGGTGAAATGCTGGAACAGTGAGCCGGAGAAGAGACCCTCCTTTTAC
     T S E V Y E I M V K C W N S E P E K R P S F Y

3037 CACCTGAGTGAGATTGTGGAGAATCTGCTGCCTGGACAATATAAAAAGAGTTATGAAAAAATTCACCTG
     H L S E I V E N L L P G Q Y K K S Y E K I H L

3106 GACTTCCTGAAGAGTGACCATCCTGCTGTGGCACGCATGCGTGTGGACTCAGACAATGCATACATTGGT
     D F L K S D H P A V A R M R V D S D N A Y I G

3175 GTCACCTACAAAAACGAGGAAGACAAGCTGAAGGACTGGGAGGGTGGTCTGGATGAGCAGAGACTGAGC
     V T Y K N E E D K L K D W E G G L D E Q R L S

3244 GCTGACAGTGGCTACATCATTCCTCTGCCTGACATTGACCCTGTCCCTGAGGAGGAGGACCTGGGCAAG
     A D S G Y I I P L P D I D P V P E E E D L G K

3313 AGGAACAGACACAGCTCGCAGACCTCTGAAGAGAGTGCCATTGAGACGGGTTCCAGCAGTTCCACCTTC
     R N R H S S Q T S E E S A I E T G S S S T F

3382 ATCAAGAGAGAGGACGAGACCATTGAAGACATCGACATGATGGACGACATCGGCATAGACTCTTCAGAC
     I K R E D E T I E D I D M M D D I G I D S S D

3451 CTGGTGGAAGACAGCTTCCTGTAACTGGCGGATTCGAGGGGTTCCTTCCACTTCTGGGGCCACCTCTGG
     L V E D S F L
                    1089

3520 ATCCCGTTCAGAAAACCACTTTATTGCAATGCGGAGGTTGAGAGGAGGACTTGGTTGATGTTTAAAGAG
3589 AAGTTCCCAGCCAAGGGCCTCGGGGAGCCGTTCTAAATATGAATGAATGGGATATTTTGAAATGAACTTT
3658 GTCAGTGTTGCCTCTTGCAATGCCTCAGTAGCATCTCAGTGGTGTGTGAAGTTTGGAGATAGATGGATA
3727 AGGGAATAATAGGCCACAGAAGGTGAACTTTGTGCTTCAAGGACATTGGTGAGAGTCCAACAGACACAA
3796 TTTATACTGCGACAGAACTTCAGCATTGTAATTATGTAAATAACTCTAACCAAGGCTGTGTTTAGATTG
3865 TATTAACTATCTTCTTTGGACTTCTGAAGAGACCACTCAATCCATCCTGTACTTCCCTCTTGAAACCTG
3934 ATGTAGCTGCTGTTGAACTTTTTAAAGAAGTGCATGAAAAACCATTTTTGAACCTTAAAAGGTACTGGT
4003 ACTATAGCATTTTGCTATCTTTTTTTAGTGTTAAAGAGATAAAGAATAATAAG

α1-1

Not I   Pst I   Not I   Xba I

5' → 3'

α1-7

Sst I   Sst I   Bam HI
Not I   Pst I   Not I

3' ← 5'

pα17R

Eco RI   Xba I   Pst I   Sst I   BamHI   Eco RI
          Not I   Not I

3' ← 5'

Eco RI   Xba I   Pst I   Sst I   Eco RI
          Not I   Not I   Cla I

3' ← 5'

α5-1

Not I   Sst I   Eco RI
                  Not I   Cla I

3' ← 5'

FIGURE 12

```
  1 GGCCCCTCAGCCCTGCTGCCCAGCACGAGCCTGTGCTCGCCCTGCCCAACGCAGACAGCCAGACCCAGG     69

 70 GCGGCCCCTCTGGCGGCTCTGCTCCTCCCGAAGGATGCTTGGGGAGTGAGGCGAAGCTGGGCGCTCCTC    138

139 TCCCCTACAGCAGCCCCCTTCCTCCATCCCTCTGTTCTCCTGAGCCTTCAGGAGCCTGCACCAGTCCTG    207

208 CCTGTCCTTCTACTCAGCTGTTACCCACTCTGGGACCAGCAGTCTTTCTGATAACTGGGAGAGGGCAGT    276

277 AAGGAGGACTTCCTGGAGGGGGTGACTGTCCAGAGCCTGGAACTGTGCCCACACCAGAAGCCATCAGCA    345

346 GCAAGGACACCATGCGGCTTCCGGGTGCGATGCCAGCTCTGGCCCTCAAAGGCGAGCTGCTGTTGCTGT    414
              M  R  L  P  G  A  M  P  A  L  A  L  K  G  E  L  L  L  L  S     20

415 CTCTCCTGTTACTTCTGGAACCACAGATCTCTCAGGGCCTGGTCGTCACACCCCCGGGGCCAGAGCTTG    483
     L  L  L  L  L  E  P  Q  I  S  Q  G  L  V  V  T  P  P  G  P  E  L  V     43

484 TCCTCAATGTCTCCAGCACCTTCGTTCTGACCTGCTCGGGGTTCAGCTCCGGTGGTGTGGGAACGGATGT    552
     L  N  V  S  S  T  F  V  L  T  C  S  G  S  A  P  V  V  W  E  R  M  S     66

553 CCCAGGAGCCCCCACAGGAAATGGCCAAGGCCCAGGATGGCACCTTCTCCAGCGTGCTCACACTGACCA    621
     Q  E  P  P  Q  E  M  A  K  A  Q  D  G  T  F  S  S  V  L  T  L  T  N     89

622 ACCTCACTGGGCTAGACACGGGAGAATACTTTTGCACCCACAATGACTCCCGTGGACTGGAGACCGATG    690
     L  T  G  L  D  T  G  E  Y  F  C  T  H  N  D  S  R  G  L  E  T  D  E    112

691 AGCGGAAACGGCTCTACATCTTTGTGCCAGATCCCACCGTGGGCTTCCTCCCTAATGATGCCGAGGAAC    759
     R  K  R  L  Y  I  F  V  P  D  P  T  V  G  F  L  P  N  D  A  E  E  L    135

760 TATTCATCTTTCTCACGGAAATAACTGAGATCACCATTCCATGCCGAGTAACAGACCCACAGCTGGTGG    828
      F  I  F  L  T  E  I  T  E  I  T  I  P  C  R  V  T  D  P  Q  L  V  V   158

829 TGACACTGCACGAGAAGAAAGGGGACGTTGCACTGCCTGTCCCCTATGATCACCAACGTGGCTTTTCTG    897
      T  L  H  E  K  K  G  D  V  A  L  P  V  P  Y  D  H  Q  R  G  F  S  G   181

898 GTATCTTTGAGGACAGAAGCTACATCTGCAAAACCACCATTGGGGACAGGGAGGTGGATTCTGATGCCT    966
      I  F  E  D  R  S  Y  I  C  K  T  T  I  G  D  R  E  V  D  S  D  A  Y   204

967 ACTATGTCTACAGACTCCAGGTGTCATCCATCAACGTCTCTGTGAACGCAGTGCAGACTGTGGTCCGCC   1035
      Y  V  Y  R  L  Q  V  S  S  I  N  V  S  V  N  A  V  Q  T  V  V  R  Q   227

1036 AGGGTGAGAACATCACCCTCATGTGCATTGTGATCGGGAATGAGGTGGTCAACTTCGAGTGGACATACC   1104
      G  E  N  I  T  L  M  C  I  V  I  G  N  E  V  V  N  F  E  W  T  Y  P   250

1105 CCCGCAAAGAAAGTGGGCGGCTGGTGGAGCCGGTGACTGACTTCCTCTTGGATATGCCTTACCACATCC   1173
      R  K  E  S  G  R  L  V  E  P  V  T  D  F  L  L  D  M  P  Y  H  I  R   273

1174 GCTCCATCCTGCACATCCCCAGTGCCGAGTTAGAAGACTCGGGGACCTACACCTGCAATGTGACGGAGA   1242
      S  I  L  H  I  P  S  A  E  L  E  D  S  G  T  Y  T  C  N  V  T  E  S   296

1243 GTGTGAATGACCATCAGGATGAAAAGGCCATCAACATCACCGTGGTTGAGAGCGGCTACGTGCGGCTCC   1311
      V  N  D  H  Q  D  E  K  A  I  N  I  T  V  V  E  S  G  Y  V  R  L  L   319
```

Fig. 1A

```
1312 TGGGAGAGGTGGGCACACTACAATTTGCTGAGCTGCATCGGAGCCGGACACTGCAGGTAGTGTTCGAGG 1380
        G  E  V  G  T  L  Q  F  A  E  L  H  R  S  R  T  L  Q  V  V  F  E  A  342

1381 CCTACCCACCGCCCACTGTCCTGTGGTTCAAAGACAACCGCACCCTGGGCGACTCCAGCGCTGGCGAAA 1449
        Y  P  P  P  T  V  L  W  F  K  D  N  R  T  L  G  D  S  S  A  G  E  I  365

1450 TCGCCCTGTCCACGCGCAACGTGTCGGAGACCCGGTATGTGTCAGAGCTGACACTGGTTCGCGTGAAGG 1518
        A  L  S  T  R  N  V  S  E  T  R  Y  V  S  E  L  T  L  V  R  V  K  V  388

1519 TGGCAGAGGCTGGCCACTACACCATGCGGGCCTTCCATGAGGATGCTGAGGTCCAGCTCTCCTTCCAGC 1587
        A  E  A  G  H  Y  T  M  R  A  F  H  E  D  A  E  V  Q  L  S  F  Q  L  411

1588 TACAGATCAATGTCCCTGTCCGAGTGCTGGAGCTAAGTGAGAGCCACCCTGACAGTGGGGAACAGACAG 1656
        Q  I  N  V  P  V  R  V  L  E  L  S  E  S  H  P  D  S  G  E  Q  T  V  434

1657 TCCGCTGTCGTGGCCGGGGCATGCCCCAGCCGAACATCATCTGGTCTGCCTGCAGAGACCTCAAAAGGT 1725
        R  C  R  G  R  G  M  P  Q  P  N  I  I  W  S  A  C  R  D  L  K  R  C  457

1726 GTCCACGTGAGCTGCCGCCCACGCTGCTGGGGAACAGTTCCGAAGAGGAGAGCCAGCTGGAGACTAACG 1794
        P  R  E  L  P  P  T  L  L  G  N  S  S  E  E  E  S  Q  L  E  T  N  V  480

1795 TGACGTACTGGGAGGAGGAGCAGGAGTTTGAGGTGGTGAGCACACTGCGTCTGCAGCACGTGGATCGGC 1863
        T  Y  W  E  E  E  Q  E  F  E  V  V  S  T  L  R  L  Q  H  V  D  R  P  503

1864 CACTGTCGGTGCGCTGCACGCTGCGCAACGCTGTGGGCCAGGACACGCAGGAGGTCATCGTGGTGCCAC 1932
        L  S  V  R  C  T  L  R  N  A  V  G  Q  D  T  Q  E  V  I  V  V  P  H  526

1933 ACTCCTTGCCCTTTAAGGTGGTGGTGATCTCAGCCATCCTGGCCCTGGTGGTGCTCACCATCATCTCCC 2001
        S  L  P  F  K  V  V  V  I  S  A  I  L  A  L  V  V  L  T  I  I  S  L  549

2002 TTATCATCCTCATCATGCTTTGGCAGAAGAAGCCACGTTACGAGATCCGATGGAAGGTGATTGAGTCTG 2070
        I  I  L  I  M  L  W  Q  K  K  P  R  Y  E  I  R  W  K  V  I  E  S  V  572

2071 TGAGCTCTGACGGCCATGAGTACATCTACGTGGACCCCATGCAGCTGCCCTATGACTCCACGTGGGAGC 2139
        S  S  D  G  H  E  Y  I  Y  V  D  P  M  Q  L  P  Y  D  S  T  W  E  L  595

2140 TGCCGCGGGACCAGCTTGTGCTGGGACGCACCCTCGGCTCTGGGGCCTTTGGGCAGGTGGTGGAGGCCA 2208
        P  R  D  Q  L  V  L  G  R  T  L  G  S  G  A  F  G  Q  V  V  E  A  T  618

2209 CGGCTCATGGCCTGAGCCATTCTCAGGCCACGATGAAAGTGGCCGTCAAGATGCTTAAATCCACAGCCC 2277
        A  H  G  L  S  H  S  Q  A  T  M  K  V  A  V  K  M  L  K  S  T  A  R  641

2278 GCAGCAGTGAGAAGCAAGCCCTTATGTCGGAGCTGAAGATCATGAGTCACCTTGGGCCCCACCTGAACG 2346
        S  S  E  K  Q  A  L  M  S  E  L  K  I  M  S  H  L  G  P  H  L  N  V  664

2347 TGGTCAACCTGTTGGGGGCCTGCACCAAAGGAGGACCCATCTATATCATCACTGAGTACTGCCGCTACG 2415
        V  N  L  L  G  A  C  T  K  G  G  P  I  Y  I  I  T  E  Y  C  R  Y  G  687

2416 GAGACCTGGTGGACTACCTGCACCGCAACAAACACACCTTCCTGCAGCACCACTCCGACAAGCGCCGCC 2484
        D  L  V  D  Y  L  H  R  N  K  H  T  F  L  Q  H  H  S  D  K  R  R  P  710

2485 CGCCCAGCGCGGAGCTCTACAGCAATGCTCTGCCCGTTGGGCTCCCCCTGCCCAGCCATGTGTCCTTGA 2553
        P  S  A  E  L  Y  S  N  A  L  P  V  G  L  P  L  P  S  H  V  S  L  T  733
```

Fig. 1B

2554 CCGGGGAGAGCGACGGTGGCTACATGGACATGAGCAAGGACGAGTCGGTGGACTATGTGCCCATGCTGG 2622
    G E S D G G Y M D M S K D E S V D Y V P M L D 756

2623 ACATGAAAGGAGACGTCAAATATGCAGACATCGAGTCCTCCAACTACATGGCCCCTTACGATAACTACG 2691
    M K G D V K Y A D I E S S N Y M A P Y D N Y V 779

2692 TTCCCTCTGCCCCTGAGAGGACCTGCCGAGCAACTTTGATCAACGAGTCTCCAGTGCTAAGCTACATGG 2760
    P S A P E R T C R A T L I N E S P V L S Y M D 802

2761 ACCTCGTGGGCTTCAGCTACCAGGTGGCCAATGGCATGGAGTTTCTGGCCTCCAAGAACTGCGTCCACA 2829
    L V G F S Y Q V A N G M E F L A S K N C V H R 825

2830 GAGACCTGGCGGCTAGGAACGTGCTCATCTGTGAAGGCAAGCTGGTCAAGATCTGTGACTTTGGCCTGG 2898
    D L A A R N V L I C E G K L V K I C D F G L A 848

2899 CTCGAGACATCATGCGGGACTCGAATTACATCTCCAAAGGCAGCACCTTTTTGCCTTTAAAGTGGATGG 2967
    R D I M R D S N Y I S K G S T F L P L K W M A 871

2968 CTCCGGAGAGCATCTTCAACAGCCTCTACACCACCCTGAGCGACGTGTGGTCCTTCGGGATCCTGCTCT 3036
    P E S I F N S L Y T T L S D V W S F G I L L W 894

3037 GGGAGATCTTCACCTTGGGTGGCACCCCCTTACCCAGAGCTGCCCATGAACGAGCAGTTCTACAATGCCA 3105
    E I F T L G G T P Y P E L P M N E Q F Y N A I 917

3106 TCAAACGGGGTTACCGCATGGCCCAGCCTGCCCATGCCTCCGACGAGATCTATGAGATCATGCAGAAGT 3174
    K R G Y R M A Q P A H A S D E I Y E I M Q K C 940

3175 GCTGGGAAGAGAAGTTTGAGATTCGGCCCCCCTTCTCCCAGCTGGTGCTGCTTCTCGAGAGACTGTTGG 3243
    W E E K F E I R P P F S Q L V L L L E R L L G 963

3244 GCGAAGGTTACAAAAAGAAGTACCAGCAGGTGGATGAGGAGTTTCTGAGGAGTGACCACCCAGCCATCC 3312
    E G Y K K K Y Q Q V D E E F L R S D H P A I L 986

3313 TTCGGTCCCAGGCCCGCTTGCCTGGGTTCCATGGCCTCCGATCTCCCCCTGGACACCAGCTCCGTCCTCT 3381
    R S Q A R L P G F H G L R S P L D T S S V L Y 1009

3382 ATACTGCCGTGCAGCCCAATGAGGGTGACAACGACTATATCATCCCCCTGCCTGACCCCAAACCCGAGG 3450
    T A V Q P N E G D N D Y I I P L P D P K P E V 1032

3451 TTGCTGACGAGGGCCCACTGGAGGGTTCCCCCAGCCTAGCCAGCTCCACCCTGAATGAAGTCAACACCT 3519
    A D E G P L E G S P S L A S S T L N E V N T S 1055

3520 CCTCAACCATCTCCTGTGACAGCCCCCTGGAGCCCCAGGACGAACCAGAGCCAGAGCCCCAGCTTGAGC 3588
    S T I S C D S P L E P Q D E P E P E P Q L E L 1078

3589 TCCAGGTGGAGCCGGAGCCAGAGCTGGAACAGTTGCCGGATTCGGGGTGCCCTGCGCCTCGGGCGGAAG 3657
    Q V E P E P E L E Q L P D S G C P A P R A E A 1101

3658 CAGAGGATAGCTTCCTGTAGGGGGCTGGCCCCTACCCTGCCCTGCCTGAAGCTCCCCCCCCTGCCAGCAC 3726
    E D S F L .   1106

3727 CCAGCATCTCCTGGCCTGGCCTGACCGGGCTTCCTGTCAGCCAGGCTGCCCTTATCAGCTGTCCCCTTC 3795

Fig. 1C

```
3796 TGGAAGCTTTCTGCTCCTGACGTGTTGTGCCCCAAACCCTGGGGCTGGCTTAGGAGGCAAGAAAACTGC 3864

3865 AGGGGCCGTGACCAGCCCTCTGCCTCCAGGGAGGCCAACTGACTCTGAGCCAGGGTTCCCCCAGGGAAC 3933

3934 TCAGTTTTCCCATATGTAAGATGGGAAAGTTAGGCTTGATGACCCAGAATCTAGGATTCTCTCCCTGGC 4002

4003 TGACACGGTGGGGAGACCGAATCCCTCCCTGGGAAGATTCTTGGAGTTACTGAGGTGGTAAATTAACAT 4071

4072 TTTTTCTGTTCAGCCAGCTACCCCTCAAGGAATCATAGCTCTCTCCTCGCACTTTTTATCCACCCAGGA 4140

4141 GCTAGGGAAGAGACCCTAGCCTCCCTGGCTGCTGGCTGAGCTAGGGCCTAGCTTGAGCAGTGTTGCCTC 4209

4210 ATCCAGAAGAAAGCCAGTCTCCTCCCTATGATGCCAGTCCCTGCGTTCCCTGGCCCGAGCTGGTCTGGG 4278

4279 GCCATTAGGCAGCCTAATTAATGCTGGAGGCTGAGCCAAGTACAGGACACCCCCAGCCTGCAGCCCTTG 4347

4348 CCCAGGGCACTTGGAGCACACGCAGCCATAGCAAGTGCCTGTGTCCCTGTCCTTCAGGCCCATCAGTCC 4416

4417 TGGGGCTTTTTTCTTTATCACCCTCAGTCTTAATCCATCCACCAGAGTCTAGAAGGCCAGACGGGCCCCG 4485

4486 CATCTGTGATGAGAATGTAAATGTGCCAGTGTGGAGTGGCCACGTGTGTGTGCCAGTATATGGCCCTGG 4554

4555 CTCTGCATTGGACCTGCTATGAGGCTTTGGAGGAATCCCTCACCCTCTCTGGGCCTCAGTTTCCCCTTC 4623

4624 AAAAAATGAATAAGTCGGACTTATTAACTCTGAGTGCCTTGCCAGCACTAACATTCTAGAGTATTCCAG 4692

4693 GTGGTGCACATTTGTCCAGATGAAGCAAGGCCTATACCCTAAACTTCATCCTGGGGGTCAGCTGGGCTC 4761

4762 CTGGGAGATTCCAGATCACACATCACACTCTGGGGACTCAGGAACCATGCCCCTTCCCCAGGCCCCCAG 4830

4831 CAAGTCTCAAGAACACAGCTGCACAGGCCTTGACTTAGAGTGACAGCCGGTGTCCTGGAAAGCCCCAAG 4899

4900 CAGCTGCCCCAGGGACATGGGAAGACCACGGGACCTCTTTCACTACCCACGATGACCTCCGGGGGTATC 4968

4969 CTGGGCAAAAGGGACAAAGAGGGCAAATGAGATCACCTCCTGCAGCCCACCACTCCAGCACCTGTGCCG 5037

5038 AGGTCTGCGTCGAAGACAGAATGGACAGTGAGGACAGTTATGTCTTGTAAAAGACAAGAAGCTTCAGAT 5106

5107 GGTACCCCAAGAAGGATGTGAGAGGTGGCTGCTTTGGAGTTTGCCCCCTCACCCCACCAGCTGCCCCAT 5175

5176 CCCTGAGGCATGCGCTCCATGGGGGTATGGTTTTGTCACTGCCCAGACCTAGCAGTGACATCTCATTGT 5244

5245 CCCCAGCCCAGTGGGCATTGGAGGTGCCAGGGGAGTCAGGGTTGTAGCCAAGACGCCCCCGCACGGGGA 5313

5314 GGGTTGGGAAGGGGGTGCAGGAAGCTCAACCCCTCTGGGCACCAACCCTGCATTGCAGGTTGGCACCTT 5382

5383 ACTTCCCTGGGATCCCCAGAGTTGGTCCAAGGAGGGAGAGTGGGTTCTCAATACGGTACCAAAGATATA 5451

5452 ATCACCTAGGTTTACAAATATTTTTAGGACTCACGTTAACTCACATTTATACAGCAGAAATGCTATTTT 5520

5521 GTATGCTGTTAAGTTTTTCTATCTGTGTACTTTTTTTTAAGGGAAAGATTTT 5572
```

Fig. 1D

Figure 2
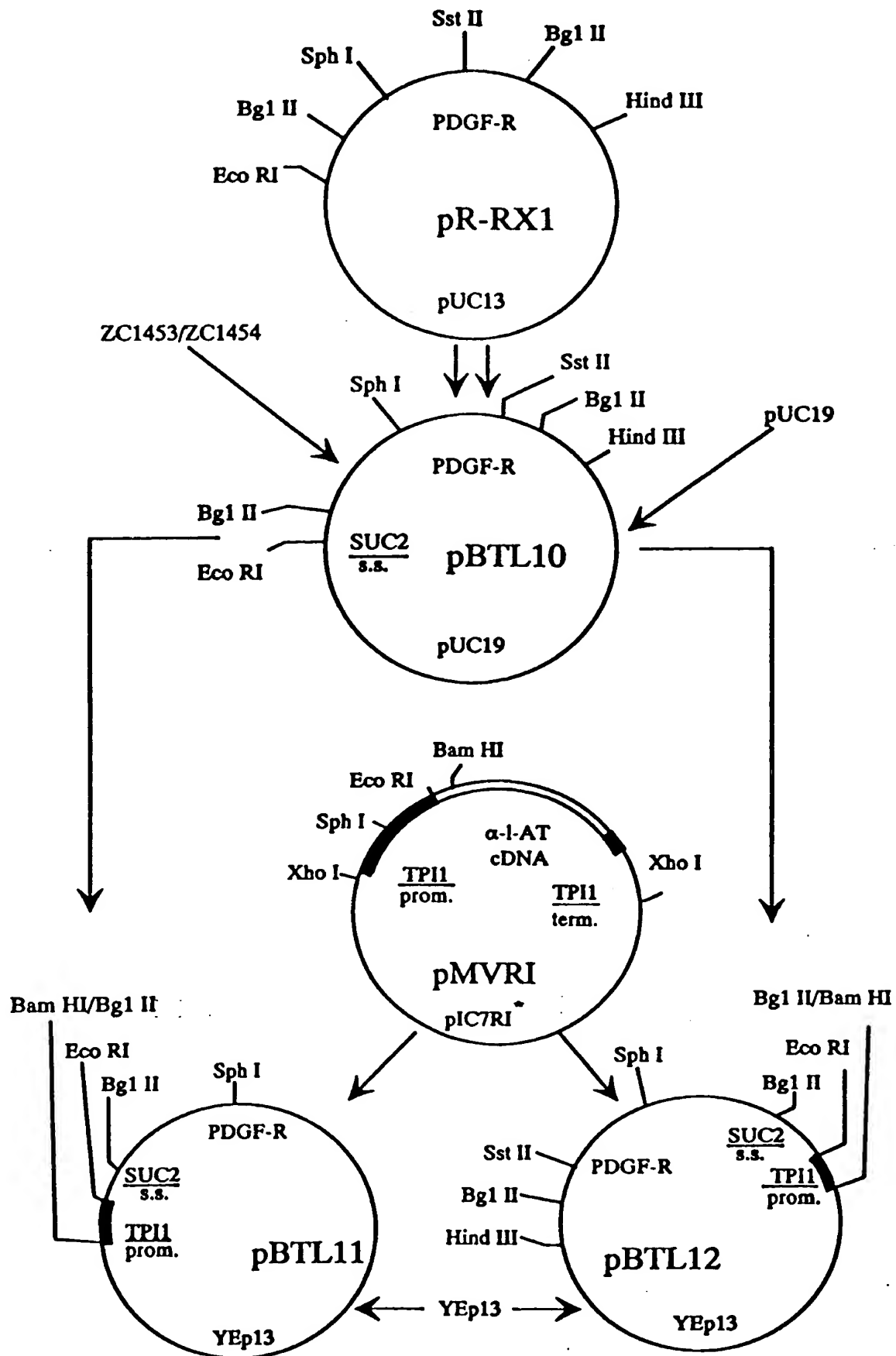
Figure 3
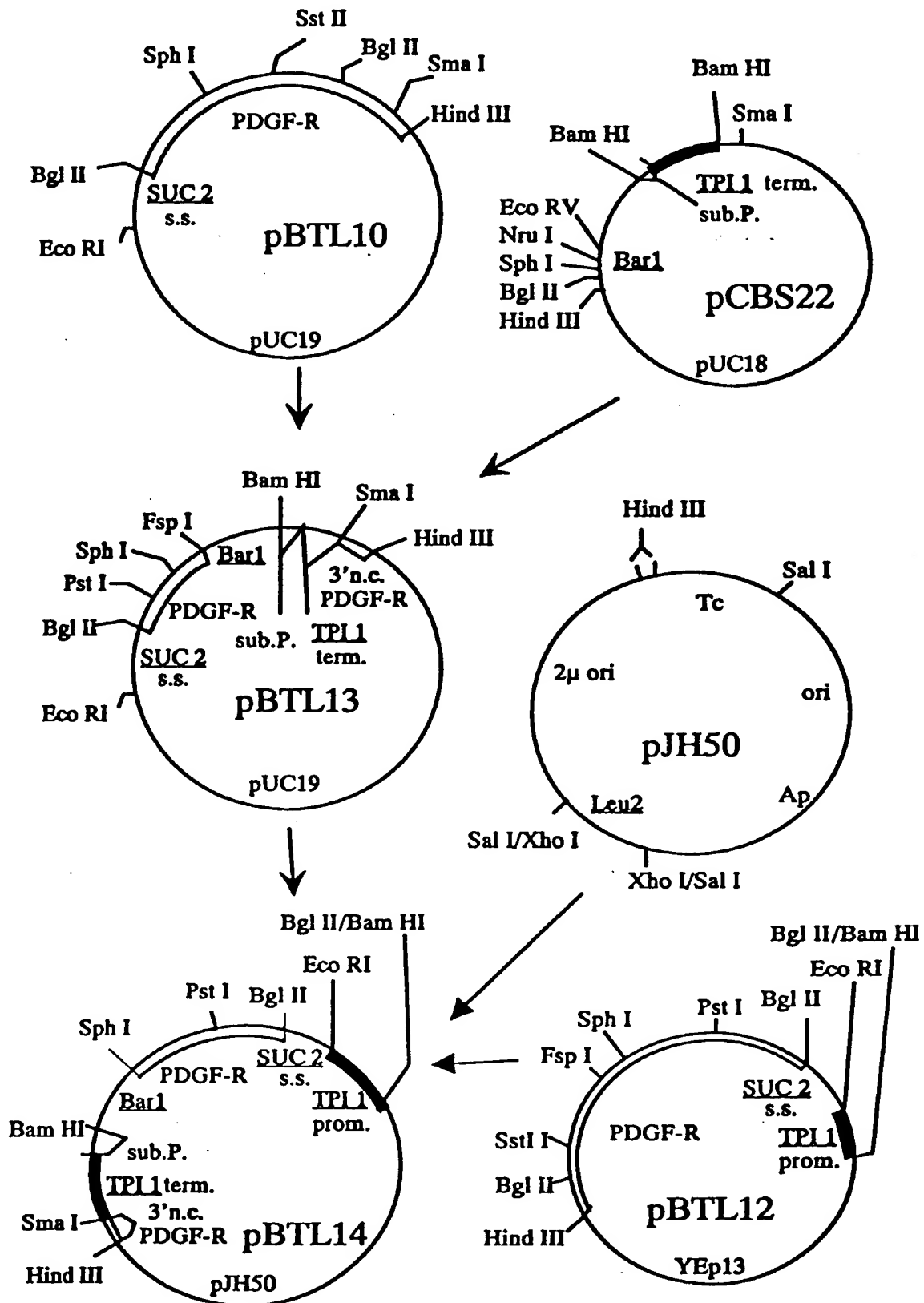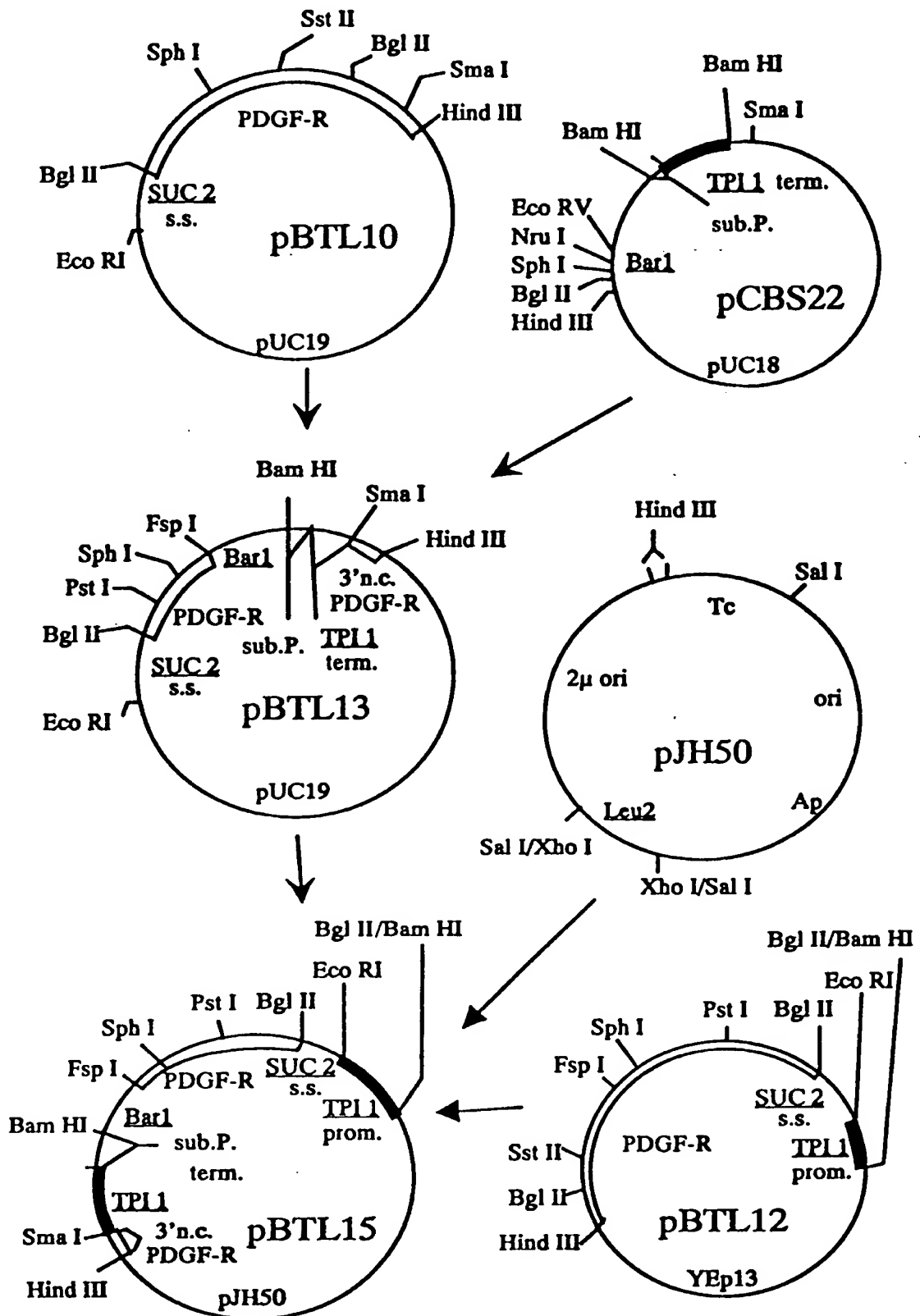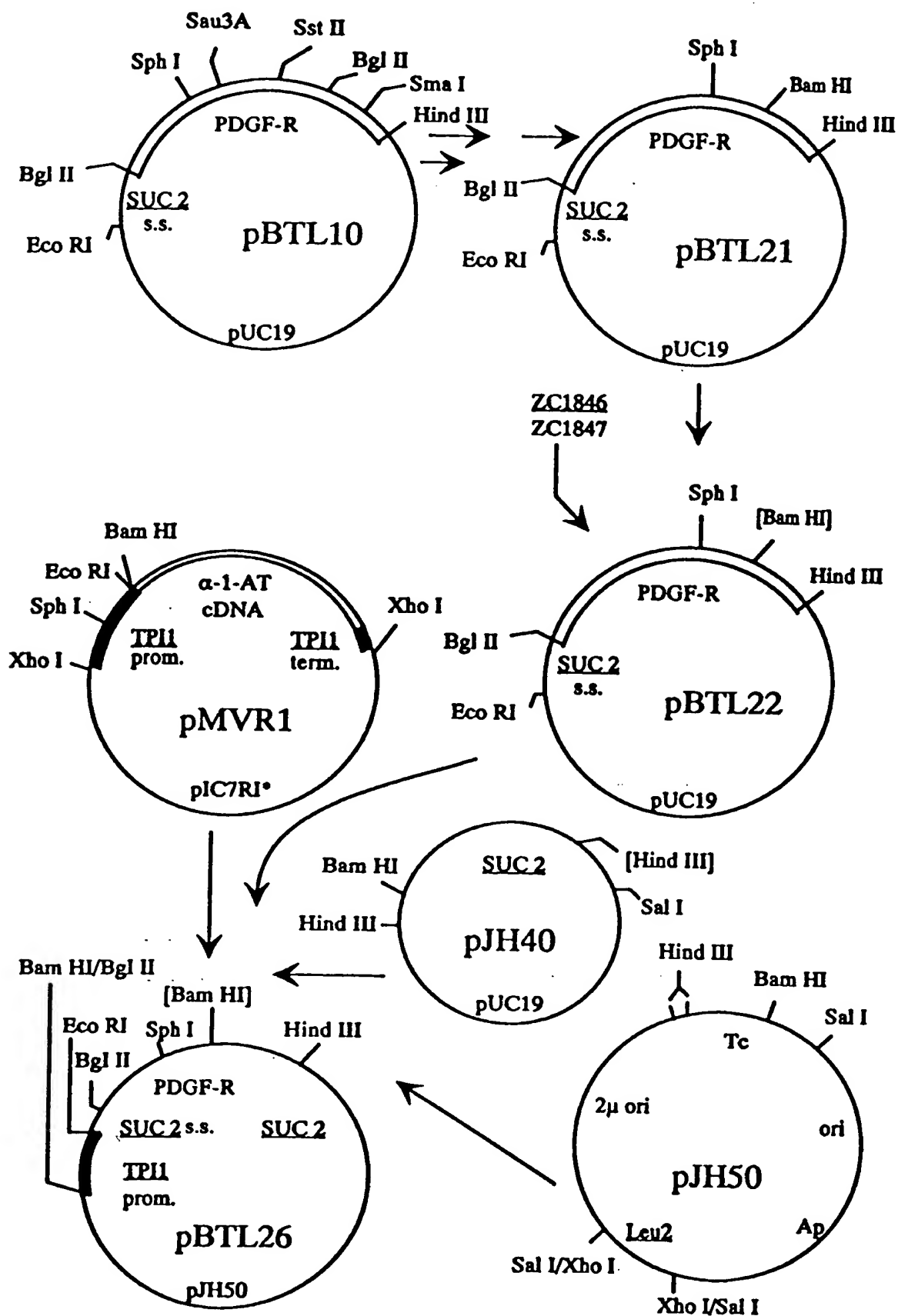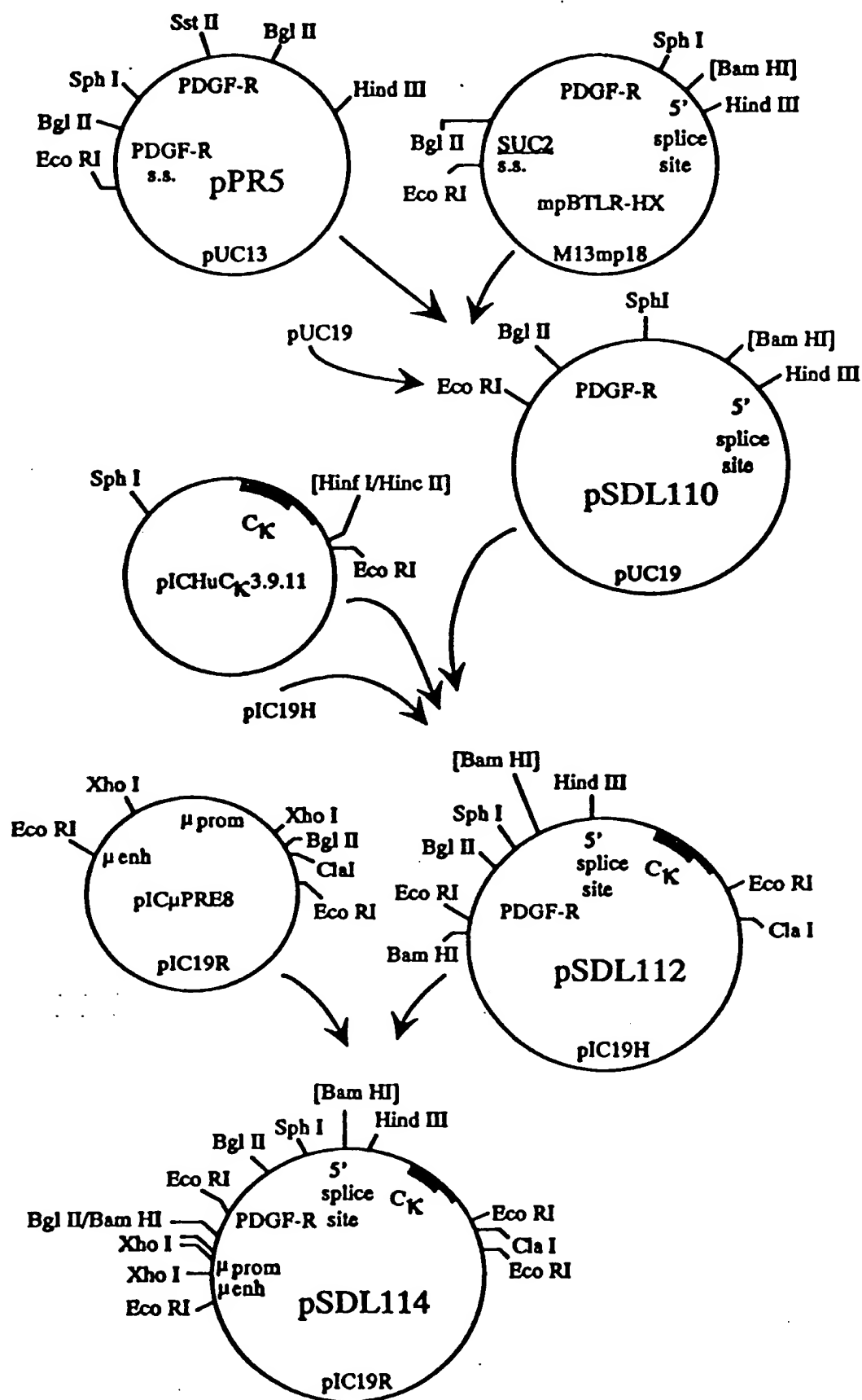
Figure 4

Figure 5
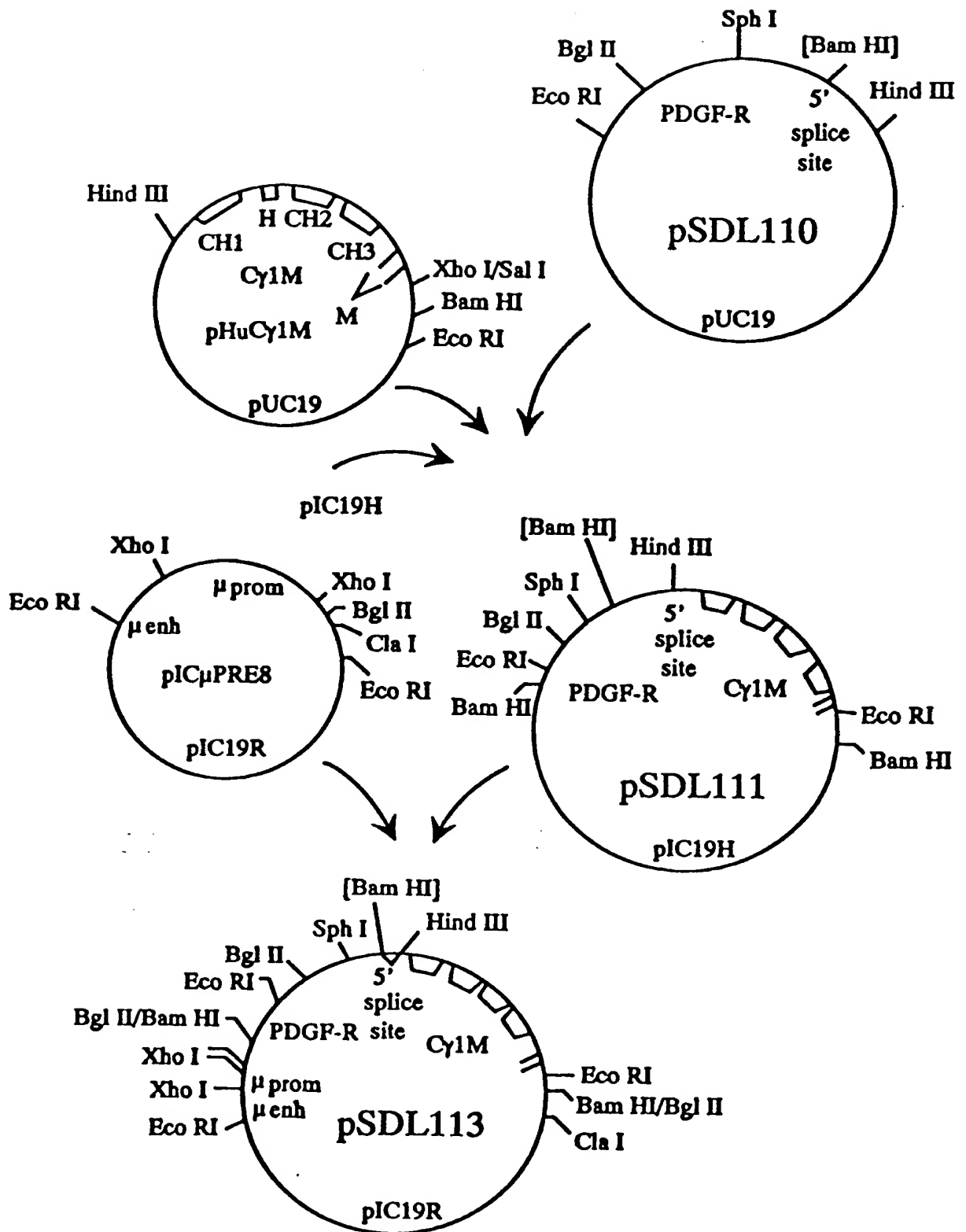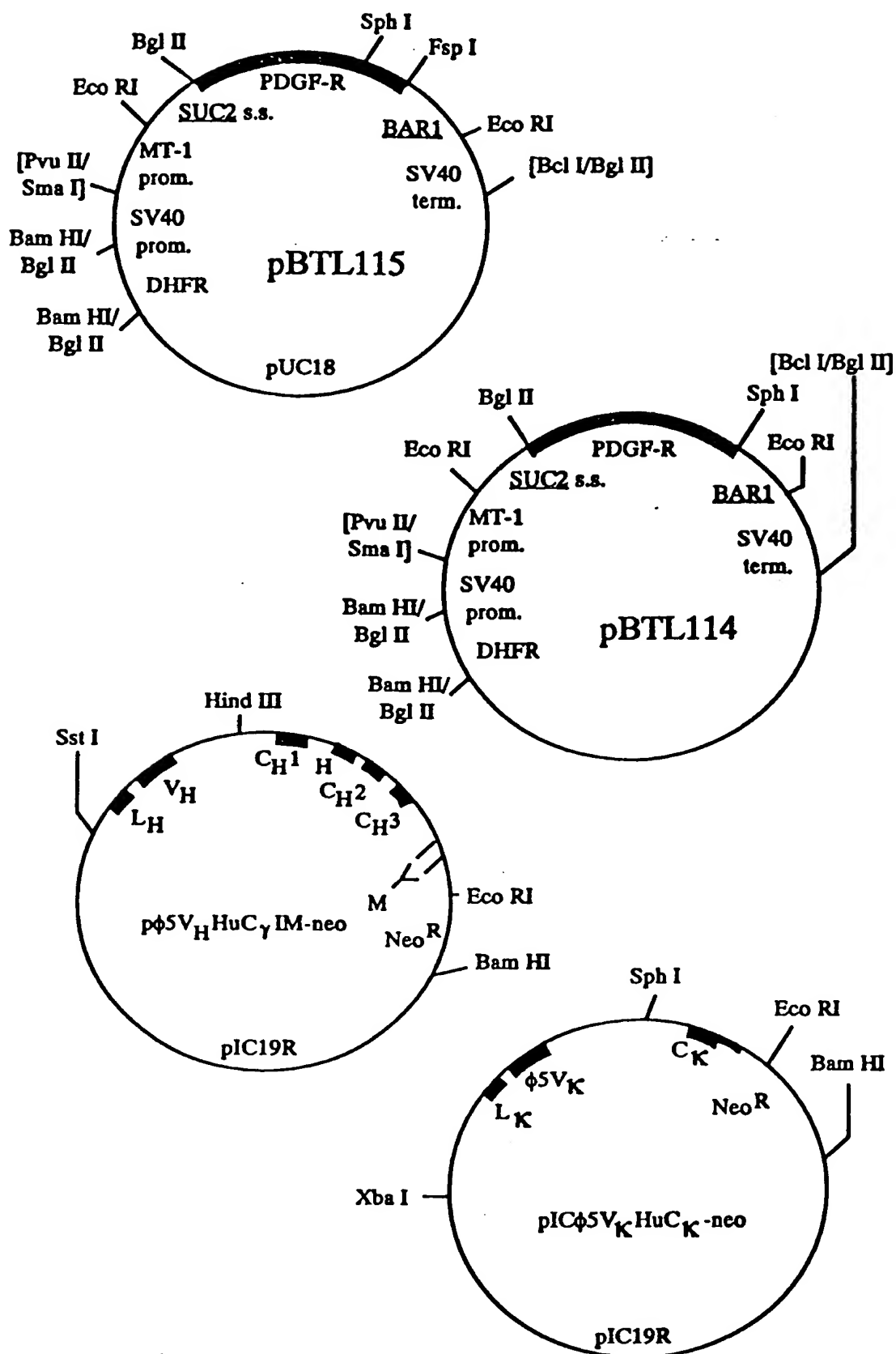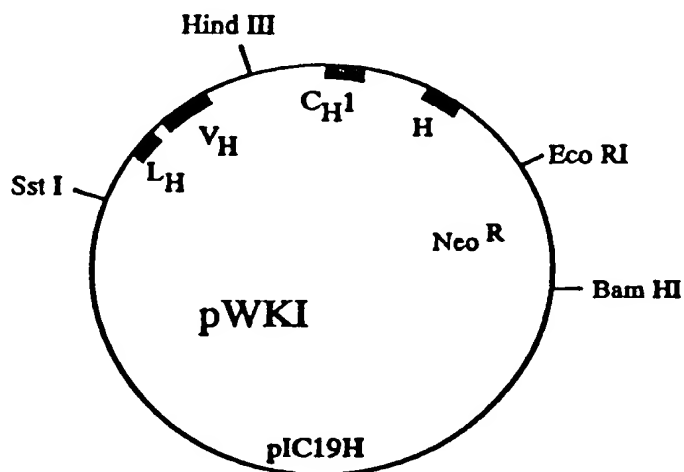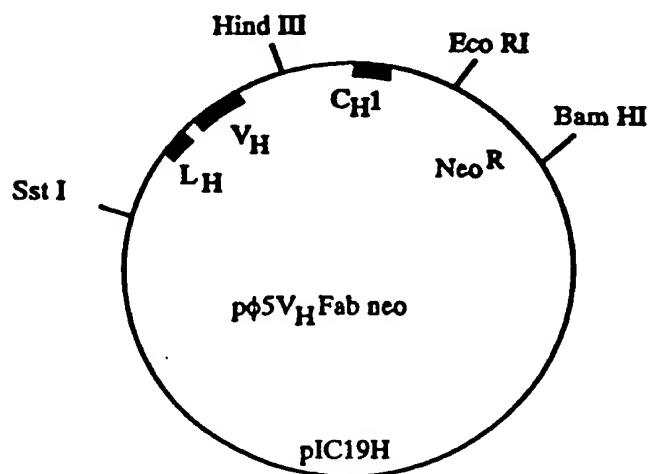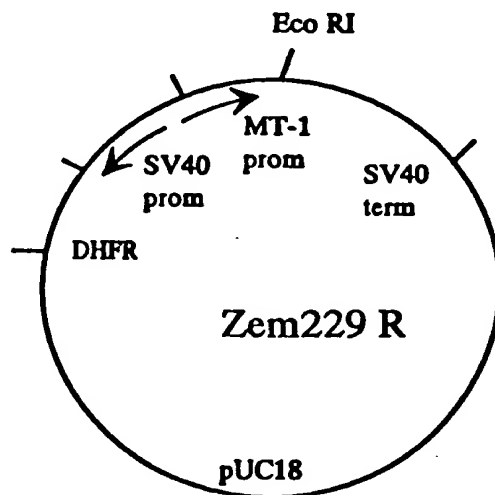
Figure 6

Figure 7

Figure 8

Figure 9

Figure 10

Figure 11A

```
  1  GCCCTGGGGACGGACCGTGGGCGGCGCGCAGCGGCGGGACGCGTTTTGGGGACGTGGTGGCCAGCGCCT
 70  TCCTGCAGACCCACAGGGAAGTACTCCCTTTGACCTCCGGGGAGCTGCGACCAGGTTATACGTTGCTGG

139  TGGAAAAGTGACAATTCTAGGAAAAGAGCTAAAAGCCGGATCGGTGACCGAAAGTTTCCCAGAGCTATG
                                                                         M
                                                                         1

208  GGGACTTCCCATCCGGCGTTCCTGGTCTTAGGCTGTCTTCTCACAGGGCTGAGCCTAATCCTCTGCCAG
      G   T   S   H   P   A   F   L   V   L   G   C   L   L   T   G   L   S   L   I   L   C   Q

277  CTTTCATTACCCTCTATCCTTCCAAATGAAAATGAAAAGGTTGTGCAGCTGAATTCATCCTTTTCTCTG
      L   S   L   P   S   I   L   P   N   E   N   E   K   V   V   Q   L   N   S   S   F   S   L

346  AGATGCTTTGGGGAGAGTGAAGTGAGCTGGCAGTACCCCATGTCTGAAGAAGAGAGCTCCGATGTGGAA
      R   C   F   G   E   S   E   V   S   W   Q   Y   P   M   S   E   E   E   S   S   D   V   E

415  ATCAGAAATGAAGAAAACAACAGCGGCCTTTTTGTGACGGTCTTGGAAGTGAGCAGTGCCTCGGCGGCC
      I   R   N   E   E   N   N   S   G   L   F   V   T   V   L   E   V   S   S   A   S   A   A

484  CACACAGGGTTGTACACTTGCTATTACAACCACACTCAGACAGAAGAGAATGAGCTTGAAGGCAGGCAC
      H   T   G   L   Y   T   C   Y   Y   N   H   T   Q   T   E   E   N   E   L   E   G   R   H

553  ATTTACATCTATGTGCCAGACCCAGATGTAGCCTTTGTACCTCTAGGAATGACGGATTATTTAGTCATC
      I   Y   I   Y   V   P   D   P   D   V   A   F   V   P   L   G   M   T   D   Y   L   V   I

622  GTGGAGGATGATGATTCTGCCATTATACCTTGTCGCACAACTGATCCCGAGACTCCTGTAACCTTACAC
      V   E   D   D   D   S   A   I   I   P   C   R   T   T   D   P   E   T   P   V   T   L   H

691  AACAGTGAGGGGGTGGTACCTGCCTCCTACGACAGCAGACAGGGCTTTAATGGGACCTTCACTGTAGGG
      N   S   E   G   V   V   P   A   S   Y   D   S   R   Q   G   F   N   G   T   F   T   V   G

760  CCCTATATCTGTGAGGCCACCGTCAAAGGAAAGAAGTTCCAGACCATCCCATTTAATGTTTATGCTTTA
      P   Y   I   C   E   A   T   V   K   G   K   K   F   Q   T   I   P   F   N   V   Y   A   L

829  AAAGCAACATCAGAGCTGGATCTAGAAATGGAAGCTCTTAAAACCGTGTATAAGTCAGGGGAAACGATT
      K   A   T   S   E   L   D   L   E   M   E   A   L   K   T   V   Y   K   S   G   E   T   I

898  GTGGTCACCTGTGCTGTTTTTAACAATGAGGTGGTTGACCTTCAATGGACTTACCCTGGAGAAGTGAAA
      V   V   T   C   A   V   F   N   N   E   V   V   D   L   Q   W   T   Y   P   G   E   V   K
```

967 GGCAAAGGCATCACAATACTGGAAGAAATCAAAGTCCCATCCATCAAATTGGTGTACACTTTGACGGTC
    G K G I T I L E E I K V P S I K L V Y T L T V

1036 CCCGAGGCCACGGTGAAAGACAGTGGAGATTACGAATGTGCTGCCCGCCAGGCTACCAGGGAGGTCAAA
    P E A T V K D S G D Y E C A A R Q A T R E V K

1105 GAAATGAAGAAAGTCACTATTTCTGTCCATGAGAAAGGTTTCATTGAAATCAAACCCACCTTCAGCCAG
    E M K K V T I S V H E K G F I E I K P T F S Q

1174 TTGGAAGCTGTCAACCTGCATGAAGTCAAACATTTTGTTGTAGAGGTGCGGGCCTACCCACCTCCCAGG
    L E A V N L H E V K H F V V E V R A Y P P P R

1243 ATATCCTGGCTGAAAAACAATCTGACTCTGATTGAAAATCTCACTGAGATCACCACTGATGTGGAAAAG
    I S W L K N N L T L I E N L T E I T T D V E K

1312 ATTCAGGAAATAAGGTATCGAAGCAAATTAAAGCTGATCCGTGCTAAGGAAGAAGACAGTGGCCATTAT
    I Q E I R Y R S K L K L I R A K E E D S G H Y

1381 ACTATTGTAGCTCAAAATGAAGATGCTGTGAAGAGCTATACTTTTGAACTGTTAACTCAAGTTCCTTCA
    T I V A Q N E D A V K S Y T F E L L T Q V P S

1450 TCCATTCTGGACTTGGTCGATGATCACCATGGCTCAACTGGGGGACAGACGGTGAGGTGCACAGCTGAA
    S I L D L V D D H H G S T G G Q T V R C T A E

1519 GGCACGCCGCTTCCTGATATTGAGTGGATGATATGCAAAGATATTAAGAAATGTAATAATGAAACTTCC
    G T P L P D I E W M I C K D I K K C N N E T S

1588 TGGACTATTTTGGCCAACAATGTCTCAAACATCATCACGGAGATCCACTCCCGAGACAGGAGTACCGTG
    W T I L A N N V S N I I T E I H S R D R S T V

1657 GAGGGCCGTGTGACTTTCGCCAAAGTGGAGGAGACCATCGCCGTGCGATGCCTGGCTAAGAATCTCCTT
    E G R V T F A K V E E T I A V R C L A K N L L

1726 GGAGCTGAGAACCGAGAGCTGAAGCTGGTGGCTCCCACCCTGCGTTCTGAACTCACGGTGGCTGCTGCA
    G A E N R E L K L V A P T L R S E L T V A A A

1795 GTCCTGGTGCTGTTGGTGATTGTGATCATCTCACTTATTGTCCTGGTTGTCATTTGGAAACAGAAACCG
    V L V L L V I V I I S L I V L V V I W K Q K P

```
1864  AGGTATGAAATTCGCTGGAGGGTCATTGAATCAATCAGCCCGGATGGACATGAATATATTTATGTGGAC
      R  Y  E  I  R  W  R  V  I  E  S  I  S  P  D  G  H  E  Y  I  Y  V  D

1933  CCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCCAAGAGATGGACTAGTGCTTGGTCGGGTCTTG
      P  M  Q  L  P  Y  D  S  R  W  E  F  P  R  D  G  L  V  L  G  R  V  L

2002  GGGTCTGGAGCGTTTGGGAAGGTGGTTGAAGGAACAGCCTATGGATTAAGCCGGTCCCAACCTGTCATG
      G  S  G  A  F  G  K  V  V  E  G  T  A  Y  G  L  S  R  S  Q  P  V  M

2071  AAAGTTGCAGTGAAGATGCTAAAACCCACGGCCAGATCCAGTGAAAAACAAGCTCTCATGTCTGAACTG
      K  V  A  V  K  M  L  K  P  T  A  R  S  S  E  K  Q  A  L  M  S  E  L

2140  AAGATAATGACTCACCTGGGGCCACATTTGAACATTGTAAACTTGCTGGGAGCCTGCACCAAGTCAGGC
      K  I  M  T  H  L  G  P  H  L  N  I  V  N  L  L  G  A  C  T  K  S  G

2209  CCCATTTACATCATCACAGAGTATTGCTTCTATGGAGATTTGGTCAACTATTTGCATAAGAATAGGGAT
      P  I  Y  I  I  T  E  Y  C  F  Y  G  D  L  V  N  Y  L  H  K  N  R  D

2278  AGCTTCCTGAGCCACCACCCAGAGAAGCCAAAGAAAGAGCTGGATATCTTTGGATTGAACCCTGCTGAT
      S  F  L  S  H  H  P  E  K  P  K  K  E  L  D  I  F  G  L  N  P  A  D

2347  GAAAGCACACGGAGCTATGTTATTTTATCTTTTGAAAACAATGGTGACTACATGGACATGAAGCAGGCT
      E  S  T  R  S  Y  V  I  L  S  F  E  N  N  G  D  Y  M  D  M  K  Q  A

2416  GATACTACACAGTATGTCCCCATGCTAGAAAGGAAAGAGGTTTCTAAATATTCCGACATCCAGAGATCA
      D  T  T  Q  Y  V  P  M  L  E  R  K  E  V  S  K  Y  S  D  I  Q  R  S

2485  CTCTATGATCGTCCAGCCTCATATAAGAAGAAATCTATGTTAGACTCAGAAGTCAAAAACCTCCTTTCA
      L  Y  D  R  P  A  S  Y  K  K  K  S  M  L  D  S  E  V  K  N  L  L  S

2554  GATGATAACTCAGAAGGCCTTACTTTATTGGATTTGTTGAGCTTCACCTATCAAGTTGCCCGAGGAATG
      D  D  N  S  E  G  L  T  L  L  D  L  L  S  F  T  Y  Q  V  A  R  G  M

2623  GAGTTTTTGGCTTCAAAAAAATTGTGTCCACCGTGATCTGGCTGCTCGCAACGTCCTCCTGGCACAAGGA
      E  F  L  A  S  K  N  C  V  H  R  D  L  A  A  R  N  V  L  L  A  Q  G

2692  AAAATTGTGAAGATCTGTGACTTTGGCCTGGCCAGAGACATCATGCATGATTCGAACTATGTGTCGAAA
      K  I  V  K  I  C  D  F  G  L  A  R  D  I  M  H  D  S  N  Y  V  S  K

2761  GGCAGTACCTTTCTGCCCGTGAAGTGGATGGCTCCTGAGAGCATCTTTGACAACCTCTACACCACACTG
      G  S  T  F  L  P  V  K  W  M  A  P  E  S  I  F  D  N  L  Y  T  T  L
```

```
2830  AGTGATGTCTGGTCTTATGGCATTCTGCTCTGGGAGATCTTTTCCCTTGGTGGCACCCCTTACCCCGGC
       S  D  V  W  S  Y  G  I  L  L  W  E  I  F  S  L  G  G  T  P  Y  P  G

2899  ATGATGGTGGATTCTACTTTCTACAATAAGATCAAGAGTGGGTACCGGATGGCCAAGCCTGACCACGCT
       M  M  V  D  S  T  F  Y  N  K  I  K  S  G  Y  R  M  A  K  P  D  H  A

2968  ACCAGTGAAGTCTACGAGATCATGGTGAAATGCTGGAACAGTGAGCCGGAGAAGAGACCCTCCTTTTAC
       T  S  E  V  Y  E  I  M  V  K  C  W  N  S  E  P  E  K  R  P  S  F  Y

3037  CACCTGAGTGAGATTGTGGAGAATCTGCTGCCTGGACAATATAAAAAGAGTTATGAAAAAATTCACCTG
       H  L  S  E  I  V  E  N  L  L  P  G  Q  Y  K  K  S  Y  E  K  I  H  L

3106  GACTTCCTGAAGAGTGACCATCCTGCTGTGGCACGCATGCGTGTGGACTCAGACAATGCATACATTGGT
       D  F  L  K  S  D  H  P  A  V  A  R  M  R  V  D  S  D  N  A  Y  I  G

3175  GTCACCTACAAAAACGAGGAAGACAAGCTGAAGGACTGGGAGGGTGGTCTGGATGAGCAGAGACTGAGC
       V  T  Y  K  N  E  E  D  K  L  K  D  W  E  G  G  L  D  E  Q  R  L  S

3244  GCTGACAGTGGCTACATCATTCCTCTGCCTGACATTGACCCTGTCCCTGAGGAGGAGGACCTGGGCAAG
       A  D  S  G  Y  I  I  P  L  P  D  I  D  P  V  P  E  E  E  D  L  G  K

3313  AGGAACAGACACAGCTCGCAGACCTCTGAAGAGAGTGCCATTGAGACGGGTTCCAGCAGTTCCACCTTC
       R  N  R  H  S  S  Q  T  S  E  E  S  A  I  E  T  G  S  S  S  S  T  F

3382  ATCAAGAGAGAGGACGAGACCATTGAAGACATCGACATGATGGACGACATCGGCATAGACTCTTCAGAC
       I  K  R  E  D  E  T  I  E  D  I  D  M  M  D  D  I  G  I  D  S  S  D

3451  CTGGTGGAAGACAGCTTCCTGTAACTGGCGGATTCGAGGGGGTTCCTTCCACTTCTGGGGCCACCTCTGG
       L  V  E  D  S  F  L
                        1089

3520  ATCCCGTTCAGAAAACCACTTTATTGCAATGCGGAGGTTGAGAGGAGGACTTGGTTGATGTTTAAAGAG
3589  AAGTTCCCAGCCAAGGGCCTCGGGGAGCGTTCTAAATATGAATGAATGGGATATTTTGAAATGAACTTT
3658  GTCAGTGTTGCCTCTTGCAATGCCTCAGTAGCATCTCAGTGGTGTGTGAAGTTTGGAGATAGATGGATA
3727  AGGGAATAATAGGCCACAGAAGGTGAACTTTGTGCTTCAAGGACATTGGTGAGAGTCCAACAGACACAA
3796  TTTATACTGCGACAGAACTTCAGCATTGTAATTATGTAAATAACTCTAACCAAGGCTGTGTTTAGATTG
3865  TATTAACTATCTTCTTTGGACTTCTGAAGAGACCACTCAATCCATCCTGTACTTCCCTCTTGAAACCTG
3934  ATGTAGCTGCTGTTGAACTTTTTAAAGAAGTGCATGAAAAACCATTTTTTGAACCTTAAAAGGTACTGGT
4003  ACTATAGCATTTTGCTATCTTTTTTTAGTGTTAAAGAGATAAAGAATAATAAG
```
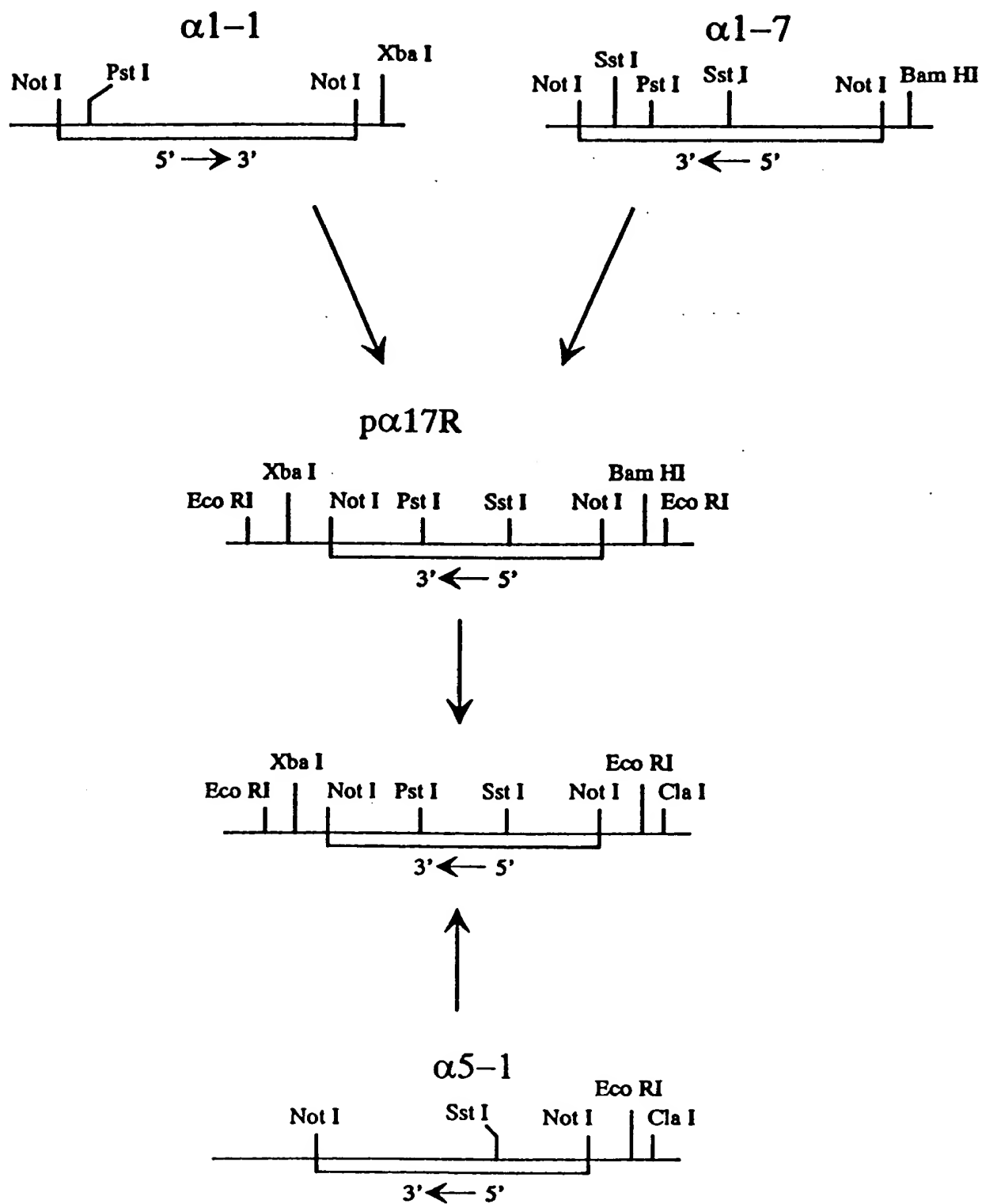
Figure 12